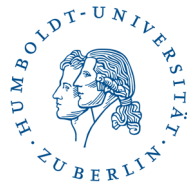


Analyse der Evaluationsdaten des Netzwerkes Schülerbefragung zur Beurteilung der Unterrichtsqualität

Diplomarbeit

zur Erlangung des akademischen Grades
eines Diplom-Volkswirt

an der Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin



vorgelegt von

Simon Deml

(Matrikel-Nr. 177800)

Erstgutachter: **Prof. Dr. W. Härdle**

Zweitgutachter: Prof. Dr. Dr. h. c. J. van Buer

Betreuer: Dr. S. Klinke

Berlin, 30. November 2009

Inhaltsverzeichnis

1	Einleitung	1
2	Datensatz	3
2.1	Hintergründe der Befragung	3
2.2	Beschreibung des Datensatzes	5
2.3	Konstrukte zur Beurteilung der Unterrichtsqualität	11
3	Statistische Methoden	14
3.1	Konfirmatorische Faktorenanalyse	14
3.2	Strukturgleichungsmodelle	25
4	Datenanalyse	33
4.1	Beobachtbarkeit der Konstrukte	33
4.2	Analyse der Konstrukte	51
5	Zusammenfassung	62
	Literaturverzeichnis	64
	Anhang A – Schülerfragebogen	68
	Anhang B – Detaillierte Item-Konstrukt-Zuordnung	70
	Anhang C – Modifizierte Item-Konstrukt-Zuordnung auf Basis der KFA-Ergebnisse	77

Abkürzungen

Abkürzung	Erklärung
ABS	Allgemein Bildende Schule
ADF	Asymptotically distribution-free (asymptotisch verteilungsfreie Schätzer)
BOS	Berufsoberschule
BS	Berufsschule
BV	Berufsvorbereitung
C.R.	Critical Ratio
CFI	Comparative-Fit-Index
d.f.	Degrees of freedom (Freiheitsgrade)
FOS	Fachoberschule
GLS	Generalized least-squares (verallgemeinerte kleinste Quadrate)
GVBI	Gesetz- und Verordnungsblatt
KFA	Konfirmatorische Faktorenanalyse
ML	Maximum-Likelihood
OBf	Berufsfachschule
OG	Berufliches Gymnasium
OSZ	Oberstufenzentrum
QuaSSU	QualitätsSicherung in Schule und Unterricht
SchulG	Schulgesetz
S.E.	Standard error (Standardfehler)
SGM	Strukturgleichungsmodell/e
SLs	Scale free least-squares (skalenunabhängige kleinste Quadrate)

SRMR	Standardized-Root-Mean-Residual
TLI	Tucker-Lewis-Index
ULS	Unweighted least-squares (ungewichtete kleinste Quadrate)

Tabellenverzeichnis

2.1	Fehlende Werte des Datensatzes	6
2.2	Häufigkeiten bei den Items „Bildungsgang“ und „Klassenstärke“	10
2.3	Kategorien des modifizierten Items „Notenschnitt“	10
2.4	Häufigkeiten beim Item „Notenschnitt“ (in Kategorien)	11
2.5	Konstrukte auf Basis der Konstruktitems	12
3.1	Kriterien und Besonderheiten der Schätzverfahren im Rahmen der KFA	20
4.1	Prüfung der multivariaten Verteilung	34
4.2	Beurteilung der Modellgüte bei Verwendung der ML-Methode [ADF-Methode]	36
4.3	Ausgewählte Faktorladungen der überidentifizierten Messmo- delle mit unzureichende Güte	39
4.4	Modifikationen der Messmodelle und deren Güte bei Verwen- dung der ML-Methode [ADF-Methode]	41
4.5	Modifikationen des Messmodells „ <i>Schülermitbestimmung</i> und <i>-autonomie</i> “ und deren Güte bei Verwendung der ML-Methode [ADF-Methode]	43
4.6	Items und Themenbereiche des Konstrukts „ <i>Schülermitbestimmung</i> und <i>-autonomie</i> “	44
4.7	Faktorladungen der gerade identifizierten, modifizierten Messmo- delle	46
4.8	Faktorladungen der gerade identifizierten, ursprünglichen Messmo- delle	49

4.9	Ergebnisse der Parameterschätzung des SGM bei Verwendung der ML-Methode [ADF-Methode] – Modell 1	54
4.10	Korrelationen mit dem Konstrukt „Interesse“	55
4.11	Ergebnisse der Parameterschätzung des SGM bei Verwendung der ML-Methode [ADF-Methode] – Modell 2	57
4.12	Korrelationen der exogenen latenten Variablen bei Verwen- dung der ML-Methode [ADF-Methode] – Modell 2	58
4.13	Ergebnisse der Parameterschätzung des SGM bei Verwendung der ML-Methode [ADF-Methode] – Modell 3	59
4.14	Ergebnisse der Parameterschätzung der Dummy-SGM bei Ver- wendung der ML-Methode	61

1 Einleitung

Im Rahmen der Diplomarbeit „Analyse der Evaluationsdaten des Netzwerkes Schülerbefragung zur Beurteilung der Unterrichtsqualität“ werden adäquate statistische Methoden im Bezug auf die Thematik der gewählten Befragung identifiziert und angewandt¹ sowie deren Ergebnisse interpretiert. Die Befragung verfolgte das originäre Ziel die Wahrnehmung der Schülerinnen und Schüler über die empfundene Qualität des Unterrichts in einer Klasse bei einem Lehrer zu messen und damit dem Lehrer oder der Lehrerin ein Feedback zu geben (Wagner 2008, S. 8). Ausgehend von den Antworten aller Schülerinnen und Schüler sollen hier durch deren statistische Auswertung mit Strukturgleichungsmodellen hauptsächlich Einflussfaktoren auf den Erfolg von Schule gemessen durch den wahrgenommenen Lernzuwachs identifiziert werden. Um diese Analyse vor allem auf Basis von Konstrukten zu ermöglichen wird zunächst durch konfirmatorische Faktorenanalysen geprüft, in wie weit es gelungen ist Konstrukte zur Abbildung von unterrichtlicher Qualität zu erfassen. Zusätzlich zu den Wahrnehmungen der Schülerinnen und Schüler werden bei der Analyse auch Merkmale wie die Klassenstärke berücksichtigt und implizit Unterschiede zwischen verschiedenen Bildungsgängen oder dem Geschlecht untersucht. Damit soll diese Arbeit dazu beitragen die interne Evaluation zu verbessern, Anhaltspunkte zur Verbesserung der Qualität von Schulunterricht liefern und Möglichkeiten der Auswertung von Befragungen dieser Art aufzeigen. Um diesen Zielen gerecht zu werden, ist bei der Interpretation der Ergebnisse darauf zu achten, dass die Meinungen der Schülerinnen und Schüler auch durch subjektive Empfindungen beeinflusst und demnach nicht alleine ausschlaggebend für die Beurteilung der Qualität von Schulunterricht sein können (Wagner 2008, S. 5/6).

¹Die Auswertung der Daten erfolgte mit Hilfe der Statistiksoftware AMOS.

Nach der Einführung in die Thematik und dem Benennen der Zielsetzung dieser Diplomarbeit werden im Datenteil (Kapitel 2) zunächst die Hintergründe und die Durchführung der Befragung beschrieben (Kapitel 2.1). Anschließend wird der zugrundegelegte Datensatz sowie dessen Besonderheiten thematisiert (Kapitel 2.2) und danach die den Ausgangspunkt bildende Annahmen bezüglich der Beziehungen zwischen den Items und unterstellten Konstrukten zur Beurteilung der Unterrichtsqualität (Kapitel 2.3) erläutert. Im Theorieteil (Kapitel 3) werden dann die zur Anwendung kommenden statistischen Verfahren diskutiert und anschließend zur Erfüllung der beschriebenen Zielsetzungen auf die Evaluationsdaten angewandt (Kapitel 4). Abschließend werden nochmals die wichtigsten Erkenntnisse und Schlussfolgerungen der durchgeführten Analysen vorgestellt (Kapitel 5).

2 Datensatz

2.1 Hintergründe der Befragung

Seit der Neufassung des Berliner Schulgesetzes im Jahr 2004 sind die Schulen in Berlin zur Qualitätssicherung und internen Evaluation ihres Bildungs- und Erziehungsauftrages verpflichtet (§9 Berliner Schulgesetz (SchulG) 2009). Ausgehend von diesem Umstand wurde auf Initiative des Oberstufenzentrums (OSZ) Druck- und Medientechnik das „Netzwerk Schülerbefragung“² gegründet und somit von einigen Schulen in Berlin ein einheitliches Verfahren zur internen Evaluation eingeführt. Durch die Zusammenarbeit wird den teilnehmenden Schulen die Durchführung von Umfragen erleichtert, da diese durch die Initiative u.a. Unterstützung beim Auswertungsprozess der Daten erhalten. Außerdem wird durch die schulübergreifende Fragebogenerhebung der Vergleich zwischen verschiedenen Schulen und somit eine bessere Einordnung der Ergebnisse ermöglicht. Seit 2005 führen die Schulen des Netzwerkes einmal jährlich Befragungen der Schülerinnen und Schüler im Bezug auf die Unterrichtsqualität durch (Wagner 2008, S. 7). Neben der Verpflichtung zur internen Evaluation wird mit der Durchführung der Befragung von Schülerinnen und Schülern aber auch im Interesse der teilnehmenden Schulen sowie Schülerinnen und Schülern gehandelt, da den Schulen und einzelnen Lehrern auf Basis der gewonnenen Informationen Ansatzpunkte zur Verbesserung der Unterrichtsqualität geliefert werden. Die Umfragen des Netzwerkes setzen dabei auf die Beurteilung des Schulunterrichts durch Schülerinnen sowie Schüler und somit durch die Empfänger von Schulbildung.

²Auch bekannt unter dem Namen „Netzwerk interne Evaluation“ (OSZ Gesundheit I 2009, S. 12).

Im Rahmen dieser Arbeit wird die Umfrage des Schuljahres 2007/08 analysiert. Im Januar 2008 nahmen 23 Berliner Schulen an der Befragung teil. Neben 18 beruflichen Schulen beteiligten sich auch zwei Gymnasien, eine Gesamtschule, eine Realschule sowie eine integrierte Haupt- und Realschule. Insgesamt wurden 21.313 Schülerinnen und Schüler aus 1.146 Klassen³ befragt (Wagner 2009a, S. 158). Die Auswahl der an der Befragung teilnehmenden Klassen einer Schule erfolgte von Schule zu Schule unterschiedlich. Während an einigen Schulen mindestens eine Klasse pro Lehrer an der Evaluierung teilnahm, wurden in anderen Schulen nicht alle Lehrer durch eine Klasse beurteilt. Auch die Auswahl einer Klasse zur Beurteilung eines Lehrers wurde unterschiedlich gehandhabt. So wurden Klassen vorgegeben, eine Zufallsauswahl durchgeführt oder in den meisten Fällen eine Klasse durch einen Lehrer ausgewählt (Wagner 2008, S. 13).

In den an der Befragung teilnehmenden Klassen wurden jeweils die einzelnen Schülerinnen und Schüler anhand des in Anhang A aufgeführten „Schülerfragebogens“ bezüglich Ihrer Wahrnehmung des Unterrichts in dieser Klasse, bei diesem Lehrer befragt. Dieser Fragebogen bildet auch die Grundlage für die in dieser Arbeit analysierten Daten. Neben dem Schülerfragebogen wurde vom Lehrer oder der Lehrerin der jeweiligen Klasse auch noch ein „Lehrerfragebogen“ ausgefüllt. Der Lehrerfragebogen wird hier dazu verwendet, um die einzelnen Schulen oder Klassen, den Bildungsgang und die Klassenstärke identifizieren zu können. Grundsätzlich enthält dieser auch noch Wahrnehmungen des Lehrers oder der Lehrerin im Bezug auf die Klasse. Da der Lehrerfragebogen nur zur Identifikation von verschiedenen Gruppen dient, wird dieser in der Arbeit nicht veranschaulicht. Der Schülerfragebogen besteht aus 88 Items. Die ersten 86 Items (im Weiteren werden diese Konstruktitems genannt, da sie grundsätzlich zur Abbildung von Konstrukten herangezogen werden) zielen auf die Unterrichtsqualität, die Selbsteinschätzung der Schülerinnen und Schüler in Bezug auf ihr unterrichtliches Engagement, ihre Leistungsfähigkeit sowie den wahrgenommenen Lernzuwachs ab (Wagner

³Durch einen bei der Befragung erhobenen Klassencode können verschiedene Klassen identifiziert werden. Im Weiteren wird auf die Besonderheiten der Erhebung noch genauer eingegangen.

2008, S. 9) und konnten mit den folgenden vier Aussagen beantwortet werden: „trifft zu“, „trifft eher zu“, „trifft eher nicht zu“ und „trifft nicht zu“. Bei dem vorletzten Item „Geschlecht“ standen die Antworten „männlich“ und „weiblich“ zur Auswahl und bei dem letzten Item „Notenschnitt“ sollte ein numerischer Wert mit einer Nachkommastelle von den Befragten eingetragen werden, der die Schulnoten des letzten Zeugnisses – also die Noten des zweiten Halbjahreszeugnisses des Schuljahres 2006/07 – im Durchschnitt widerspiegelt sollte. Der Schülerfragebogen des Netzwerkes Schülerbefragung wurde in Anlehnung an die QuaSSU-Studie⁴ von Ditton und Merz (2001) sowie die Schülerbefragung in beruflichen Schulen von Seeber und Squarra (2003) konstruiert und demnach können durch die Konstruktitems 20 verschiedene Konstrukte nachgebildet werden (Wagner 2009a, S. 159).

2.2 Beschreibung des Datensatzes

Der Datensatz der Befragung zur Unterrichtsqualität enthält nicht alle Urteile der 21.313 befragten Schülerinnen und Schülern. Auf Grund des Ausschlusses von Fragebögen ohne jegliche Angaben, Fragebögen mit mehr als 75% fehlenden Antworten bzw. Werten, Fragebögen mit mehr als 77 identischen Antworten und Fragebögen mit eindeutigem Antwortmuster reduziert sich die Anzahl der im Datensatz enthaltenen Fragebögen auf 20.481.⁵ Die Gründe für den Ausschluss dieser Schülerfragebögen sind hauptsächlich, dass in einigen Fällen keine Bereitschaft zur Teilnahme an der Befragung unterstellt werden musste und dass das Angeben von identischen Antworten auf Grund von entgegengesetzt gepolten Items semantisch nur begrenzt korrekt sein kann (Wagner 2008, S. 13). In einem weiteren Schritt wurden noch vereinzelte fehlende Werte aus den 86 Konstruktitems des Datensatzes mit 20.481 Schülerinnen und Schülern ersetzt, um die angestrebten statistischen Analy-

⁴Weitere Informationen zu QualitätsSicherung in Schule und Unterricht (QuaSSU) können auf der Projekthomepage eingesehen werden (www.quassu.net).

⁵Im Datensatz sind außerdem statt der 1.146 befragten Klassen nur 1.143 Klassen ausgewiesen.

severfahren einsetzen zu können. Ersetzt wurden die vereinzelt, fehlenden Werte von Items auf Basis der Antworten der drei Items mit den ähnlichsten Aussagen⁶ (Wagner 2008, S. 16). Die Anzahl der vereinzelt fehlenden Werte lag unter 5% und diese waren unsystematisch verteilt (Wagner 2008, S. 13). Der zu untersuchende Datensatz lag in der um fehlende Werte bereinigten Form vor, weshalb hier keine weitere Auseinandersetzung mit dem Umgang von fehlenden Werten wie Beibehaltung, Ausschluss oder Imputation für die 86 Konstruktitens durchgeführt wird.

Der wie im vorhergehenden Abschnitt beschriebene Datensatz mit den 88 Items des Schülerfragebogens und den dazugehörigen 20.481 Antworten der Schülerinnen und Schüler bildet den Ausgangspunkt für diese Arbeit. Zusätzlich zu den 88 Items wurden dem Datensatz noch die Items „Schulcode“, „Klassencode“, „Bildungsgang“ und „Klassenstärke“ hinzugefügt. Diese Items wurden mit Hilfe des im Vorhergehenden erwähnten Lehrerfragebogens erhoben. Beim Item Bildungsgang konnten die Antworten „Klasse 7 & 8 – Allgemein bildende Schule (ABS)“, „Klasse 9 & 10 – ABS“, „Klasse 11 - 13 – ABS“, „Berufsvorbereitung (BV)“, „Berufsfachschule (OBF)“, „Berufsschule (dual; BS)“ sowie „Fachoberschule/Berufsoberschule/Berufliches Gymnasium (FOS/BOS/OG)“ gegeben werden und beim Item Klassenstärke standen die Antworten „bis 15 Schüler“, „16 - 20 Schüler“, „21 - 25 Schüler“ sowie „über 25 Schüler“ zur Auswahl. Im Gegensatz zu den 86 Konstruktitens sind bei den verbleibenden Items weiterhin fehlende Werte im Datensatz. In Tabelle 2.1 sind die Items mit fehlenden Werten im Datensatz und deren Häufigkeiten veranschaulicht.

Tab. 2.1: Fehlende Werte des Datensatzes

Item	Anzahl der fehlenden Werte	in Prozent
Geschlecht	1.602	7,8
Notenschnitt	3.159	15,4
Bildungsgang	3.271	16,0
Klassenstärke	2.686	13,1

⁶Durch Reliabilitätsprüfungen konnten keine gravierenden Veränderungen durch diese Art der Ersetzung von fehlenden Werten nachgewiesen werden.

Eine Besonderheit dieser fehlenden Werte ist, dass die Bereitschaft der Beantwortung der Frage zum Notenschnitt deutlich geringer war als bei anderen Fragen. Ein Grund dafür könnte auch sein, dass bei Beantwortung des Items „Notenschnitt“ die Anonymität der Befragten gegenüber deren Lehrer nicht mehr in vollem Umfang gewährleistet ist. Dies würde insbesondere auf Schülerinnen und Schüler mit sehr guten oder sehr schlechten Noten zutreffen. Weiterhin ist bei den Items Bildungsgang und Klassenstärke der prozentuale Anteil an fehlenden Werten von deutlich über zehn Prozent vor allem darauf zurückzuführen, dass die Schulen mit den meisten Befragten keine Lehrerbefragung durchführten. Bei diesen beiden Items betreffen die fehlenden Werte außerdem immer ganze Klassen, da diese Items wie bereits erwähnt nur einmal je Klasse erhoben wurden (Wagner 2008, S. 13). Die fehlenden Werte dieser Items werden bei den folgenden Untersuchungen jeweils als zusätzliche Kategorie mit der Bezeichnung „Unbekannt“ einbezogen.

Die Items „Geschlecht“, „Bildungsgang“, „Klassenstärke“ sowie der „Schulcode“ und der „Klassencode“ sind nominal skaliert und fließen im Rahmen der Strukturgleichungsmodelle (SGM) als Dummy-Variablen⁷ ein. Das Item „Notenschnitt“ ist ordinalskaliert, da eine Rangordnung mit ungleichen Abständen vorhanden ist. Auf die Besonderheiten bei diesem Item wird noch gesondert eingegangen. Hingegen wird den 86 Konstruktitems eine diskrete Intervallskalierung unterstellt und somit die Möglichkeit eröffnet, die angestrebten Analysemethoden anwenden zu können. Demnach wurden den vier Antwortmöglichkeiten dieser Items die Werte eins bis vier gegeben. Der Antwortmöglichkeit „trifft nicht zu“, die die Ablehnung einer Aussage signalisiert, wurde der Wert eins zugeordnet. Mit abnehmender Ablehnung einer Aussage bzw. mit zunehmender Zustimmung zu einer Aussage wurden den Antwortmöglichkeiten „trifft eher nicht zu“, „trifft eher zu“ und „trifft zu“

⁷Normalerweise können nur metrisch skalierte Variablen in ein Strukturgleichungsmodell einfließen. Transformiert man allerdings nominale Variablen in Dummy-Variablen, d.h. binäre Variablen mit den Werten null oder eins, können diese Variablen auch in einem Strukturgleichungsmodell berücksichtigt werden. Mit dem Wert eins wird beispielsweise bei der Dummy-Variablen „Geschlecht“ ausgedrückt, dass es sich bei einer befragten Person um eine männliche Person handelt, und mit dem Wert null, dass es sich um keine männliche Person bzw. eine weibliche Person handelt (Backhaus et al. 2006, S. 9).

dementsprechend die Werte zwei, drei und vier zugeordnet und damit eine Entscheidung bezüglich der Metrik getroffen. Bei den Items „ud04“, „mb04“, „mc02“ und „md02“ wurden den Antwortmöglichkeiten die Werte in umgekehrter Reihenfolge zugeordnet und somit die Items recodiert, damit inhaltlich gleichgerichtete Antworten von Items eines bestimmten Sachverhaltes bei den verschiedenen Analyseschritten untersucht werden.⁸ Im Weiteren werden die Bezeichnungen der recodierten Items um ein „r“ erweitert und damit die Recodierung kenntlich gemacht.

Im Gegensatz zu nominal- oder ordinalskalierten Items können bei intervallskalierte Items grundsätzlich Extremwerte auftreten. Da die Werte der Konstruktitens jedoch auf die Zahlen eins bis vier begrenzt sind, ist ein Ausschluss oder eine Ersetzung nicht sinnvoll. Außerdem muss der Umgang mit Extremwerten auch im Kontext der Thematik gesehen werden. Würde man beispielsweise vorhandene Extremwerte aussortieren, würde man abweichende Meinungen bezüglich der empfundenen Qualität von Unterricht nur zu einem bestimmten Maß zulassen und könnte damit u.U. nicht mehr von einem vollständigen Meinungsbild ausgehen.

Die Anzahl der Befragten Schülerinnen und Schüler je Schule oder Klasse ist sehr verschieden. Die kleinste Zahl an Befragten je Schule ist 161 und die größte Zahl ist 2.227. Acht Schulen haben höchstens 500 Befragte, sechs Schulen haben zwischen 501 und 1.000 Befragte, weitere sechs Schulen haben zwischen 1.001 und 1.500 Befragte. Die verbleibenden drei Schulen haben mehr als 1.500 befragte Schülerinnen und Schüler. Bei der Anzahl der Befragten je Klasse ist der niedrigste Wert drei und der höchste Wert 75 Schülerinnen und Schüler pro Klasse. Basierend auf den Klassencodes, anhand derer man die Anzahl der Befragten pro Klasse identifizieren kann, haben 32 Klassen mehr als 34 befragte Schülerinnen und Schüler (2,8 Prozent der Klassen, 7,9 Prozent aller Befragten). Eine Anzahl an befragten Schülern pro Klasse von 34 bis zu 75 scheint nur eingeschränkt plausibel. Allerdings ist eine Plausibilitätsprüfung zusammen mit dem Item „Klassenstärke“ nicht möglich, da

⁸Die Recodierung erfolgte auf Basis von sachlogischen Überlegungen und vorhergehender Untersuchungen durch Wagner (2008).

es sich bei diesen Klassen hauptsächlich um Klassen von der Schulen ohne abgegebene Lehrerfragebögen handelt. Wahrscheinlich wurde bei diesen Klassencodes eine Lehrerin oder ein Lehrer von mehreren Klassen ohne Verwendung eines separaten Klassencodes beurteilt (Wagner 2008, S. 13). Ein Vergleich zwischen der Anzahl der Befragten je Klasse, in den Kategorien des Items „Klassenstärke“, und der Klassenstärke ergab, dass bei elf Klassen (1,2% der Befragten bzw. 1,0% der verschiedenen Klassencodes) die Anzahl der Befragten größer ist als die Klassenstärke und dass bei 97 Klassen (7,6% der Befragten bzw. 8,5% der Klassen) die Anzahl der Befragten je Klasse um mindestens zwei Kategorie kleiner ist als die Klassenstärke. Zum Einen ist daher ein Teil der Werte beim Item „Klassenstärke“ nicht plausibel und zum Anderen könnte die Beteiligung an der Befragung bzw. die Rücklaufquote z.T. verbessert werden. Bei 457 Klassen (40,0% der Klassencodes) ist die Anzahl der Befragten nur um eine Kategorie kleiner als die Klassenstärke, bei 491 Klassen (43,0% der Klassencodes) ist die Anzahl der Befragten in derselben Kategorie wie die Klassenstärke und zu den restlichen Klassen können keine Aussagen getroffen werden, da die Klassenstärke unbekannt ist. Insgesamt kann man von einer zufriedenstellenden Beteiligung an der Befragung ausgehen. Weiterhin spielen die Datenungereimtheiten bei der Anzahl der Befragten pro Klasse und dem Item Klassenstärke nur eine untergeordnete Rolle im Zusammenhang mit dieser Arbeit, da sie auf Basis aller Wahrnehmungen der Schülerinnen und Schüler Ansatzpunkt zur Verbesserung des Lernerfolgs gemessen am Lernzuwachs liefern soll und der Einfluss der Anzahl der befragten Schülerinnen und Schüler pro Klasse und die Klassenstärke darauf nur einen kleinen, zusätzlichen Teilaspekt dieser Arbeit darstellt. Bei der Interpretation der Analyseergebnisse müssen diese Erkenntnisse aber beachtet werden.

Grundsätzlich besteht der Datensatz aus Wahrnehmungen von mehr Schülerinnen (51,1%) als Schülern (41,1%) und bei 7,8 Prozent der Befragten ist das Geschlecht unbekannt. Weiterhin sind hauptsächlich berufliche Schulen in der Stichprobe vertreten und das Merkmal mit den meisten Ausprägungen beim Bildungsgang ist die „Berufsschule“ (BS; siehe Tab. 2.2). Beim Item

Klassenstärke ist das häufigste Merkmal eine Klassenstärke „zwischen 12 und 25 Schülern“ und das seltenste Merkmal eine Klassenstärke von „bis zu 15 Schülern“ (siehe Tab 2.2.). Die Problematik von unbekannten bzw. fehlenden Werten wurde bereits thematisiert.

Tab. 2.2: Häufigkeiten bei den Items „Bildungsgang“ und „Klassenstärke“

Item	Ausprägung	Häufigkeit (in %)
Bildungsgang	Klasse 7 & 8 – ABS	2,4
	Klasse 9 & 10 – ABS	4,7
	Klasse 11 - 13 – ABS	5,7
	BV	3,0
	OBF	18,2
	BS	32,3
	FOS/BOS/OG	17,7
	Unbekannt	16,0
Klassenstärke	bis 15 Schüler	12,1
	16- 20 Schüler	18,6
	21 - 25 Schüler	30,5
	über 25 Schüler	25,7
	Unbekannt	13,1

Die Antworten beim Item „Notenschnitt“ sind sehr heterogen und liegen zwischen 1,0 und 6,0. Diese Werte spiegeln die durchschnittlichen Schulnoten der befragten Schülerinnen und Schüler des Zeugnisses im Sommer 2007 wieder. Allerdings wurde die Befragung ca. sechs Monate später durchgeführt und es ist deshalb nicht davon auszugehen, dass sich alle Schülerinnen und Schüler an den genauen Notenschnitt erinnern konnten. Aus diesem Grund werden

Tab. 2.3: Kategorien des modifizierten Items „Notenschnitt“

Kategorie	Werte der Kategorie
1 (sehr gut)	1,0 - 1,4
2 (gut)	1,5 - 2,4
3 (befriedigend)	2,5 - 3,4
4 (ausreichend)	3,5 - 4,4
5 (mangelhaft)	4,5 - 5,4
6 (ungenügend)	5,5 - 6,0

die erhobenen, zuordbaren Werte wie in Tabelle 2.3 ersichtlich in Kategorien unterteilt und auf dieser Basis ausgewertet.

Weiterhin bezieht sich der Notenschnitt auf das gesamte Zeugnis und nicht nur auf den Unterricht eines Lehrers, bei dem die Befragung durchgeführt wurde. Aus diesem Grund wurde entschieden, dass das Item „Notenschnitt“ nicht wie das Konstrukt „Lernzuwachs“ als abhängige Variable zur Identifikation von Einflussfaktoren auf den Erfolg von Schule sondern als zusätzliche Variable in Dummy-Form genutzt wird. In Tab. 2.4 sind die Häufigkeiten der Ausprägungen des modifizierten Items „Notenschnitt“ abgebildet.

Tab. 2.4: Häufigkeiten beim Item „Notenschnitt“ (in Kategorien)

Kategorie	Häufigkeit (in %)
Sehr gut	4,2
Gut	28,4
Befriedigend	40,2
Ausreichend	10,2
Mangelhaft	1,3
Ungenügend	0,2
Unbekannt	15,4

2.3 Konstrukte zur Beurteilung der Unterrichtsqualität

Wie bereits in Abschnitt 2.1 erwähnt wurde der Schülerfragebogen mit 86 Konstruktitems ursprünglich gebildet, um 20 Konstrukte aus sechs Konstruktgruppen abbilden zu können. Im Vorfeld dieser Arbeit wurden durch Wagner (2008) bereits Auswertungen der Daten durchgeführt. Neben der Korrelationsanalyse stellte dabei auch die Reliabilitätsanalyse zur Beurteilung der inneren Konsistenz von Items eines Konstrukts mit Hilfe des Cronbach- α -Konsistenzkoeffizienten⁹ einen Aspekt der Auswertung dar. Basierend

⁹ $\alpha = \frac{c}{c-1} \cdot \left(1 - \frac{\sum_{i=1}^j S_i^2}{S_x^2}\right)$, mit S_i^2 = Varianz des Testitems, c = Anzahl der des Testitems, S_x^2 = Varianz des Gesamtwerts der Skala des Testitems (Bühner 2006, S. 132).

auf diesen Auswertungen wird die zu untersuchende Zuordnung von Items zu Konstrukten festgelegt. Statt aller 86 Konstruktitems werden nur 80 Items zur Abbildung der in Tabelle 2.5 beschriebenen Konstrukte verwendet.

Tab. 2.5: Konstrukte auf Basis der Konstruktitems

Konstruktgruppe	unterstelltes Konstrukt	Items des Konstrukts
Lehrerpersönlichkeit	<ul style="list-style-type: none"> • <i>Fachkompetenz</i> • <i>Unterrichtsführung</i> und <i>Strukturiertheit</i> • Diagnostische Kompetenz (<i>Diagnostik</i>) • <i>Lehrer-Schüler-Sozialklima</i> 	ui01 - ui03 ua01 - ua07 ug01 - ug03 uh01 - uh04
Unterrichtsangebot	<ul style="list-style-type: none"> • <i>Veranschaulichung</i> • <i>Interessantheit</i> • <i>Transparenz und Feedback</i> • <i>Unterstützende Materialien</i> • <i>Schüleraktivierender Unterricht (Schüleraktivierung)</i> • <i>Schülermitbestimmung und -autonomie</i> • <i>Unangemessenheit der Mindestanforderungen</i> • <i>Unangemessenheit der Leistungserwartungen</i> • <i>Binnendifferenzierung</i> • <i>Klassen- und Zeitmanagement</i> 	ub01 - ub05 uc01 - uc04 ma01 - ma04 la01 - la04 ld01 - ld03, ld05 le01 - le05, lf01 - lf03, lf05 uf01 - uf03 ue01 - ue03 lb01 - lb03 ud01 - ud05
Unterrichtsnutzung	<ul style="list-style-type: none"> • <i>Unterrichtliches Selbstengagement</i> 	mc01 - mc04
Schülervoraussetzungen	<ul style="list-style-type: none"> • <i>Selbsteinschätzung zur Leistungsfähigkeit</i> • <i>Fachbezogenes Interesse</i> 	mb01 - mb04r md01 - md04
Kontext	<ul style="list-style-type: none"> • <i>Einstellung zur Schule (Schulbeurteilung)</i> 	ga01 - ga03
Unterrichtswirksamkeit	<ul style="list-style-type: none"> • <i>Wahrgenommener Lernzuwachs</i> 	za01 - za04

So wird das Item „ld04“ nicht verwendet, da es laut Wagner (2009b) nicht hinreichend mit dem Konstrukt „Schüleraktivierender Unterricht“ korrelierte¹⁰ und das Item „lf04“, weil es auf Grund von organisatorischen Fehlern

¹⁰Dies bestätigte sich auch bei einer Überprüfung mit Hilfe der KFA. Auch bei den anderen strittigen Items wurden KFAs zur Überprüfung durchgeführt. Thematisiert werden diese hier allerdings nicht, da sie nur die vorherigen Ergebnisse bestätigten.

nicht auf allen Fragebögen zu finden war. Außerdem konnten die Items „lc01“ bis „lc04“ auf Grund mangelnder innerer Konsistenz nicht zur Bildung eines Konstrukts verwendet werden. Somit werden die Analysen dieser Arbeit auf Basis von 80 Konstruktitems und 19 angenommenen, durch die Items abgebildeten Konstrukte durchgeführt. Die Bezeichnungen der unterstellten Konstrukte wurden in leicht modifizierter Form von Wagner (2008) übernommen und die kursiv geschriebenen Teile der unterstellten Konstrukte sind die im Weiteren verwendeten, verkürzten Bezeichnungen für die Konstrukte. Die Bezeichnungen der Konstruktgruppen wurden in Anlehnung an das „Angebots-Nutzungs-Modell unterrichtlicher Wirkung“ von Helmke (2003) gewählt und dienen hier vor allem dazu, die Vergleichbarkeit mit vorherigen Auswertungen der Daten des Netzwerkes Schülerbefragung für die Schulen, die an der Befragung teilgenommen haben, zu erhöhen. Gleichzeitig wird dadurch die Übersichtlichkeit der 19 Konstrukte verbessert. Im Anhang B sind nochmals alle unterstellten Konstrukte und die dazugehörigen, verwendeten Items dargestellt, um die Verständlichkeit im Bezug auf die Aussagen der einzelnen Konstrukte zu erhöhen und um bei Bedarf die Überprüfung der Interpretationen der Ergebnisse zu erleichtern. Auf Konstruktebene bedeuten Werte zwischen 1 und 2,4 eine ablehnende Haltung, der Wert 2,5 eine indifferente Haltung und Werte zwischen 2,6 und 4 eine zustimmende Haltung bezüglich des hypothetischen Konstrukts.¹¹ So steht beispielsweise ein hoher Wert beim Konstrukt „Interessantheit“ dafür, dass eine Schülerin bzw. ein Schüler den Unterricht bei seinem Lehrer als interessant empfindet. Dementsprechend steht ein niedriger Wert für einen uninteressant empfundenen Unterricht.

¹¹Dies gilt, da die Metrik der latenten Variablen im Rahmen der KFA und der SGM analog zur Metrik der Items bzw. Indikatorvariablen festgelegt wird.

3 Statistische Methoden

3.1 Konfirmatorische Faktorenanalyse

Die konfirmatorische Faktorenanalyse (KFA) ist ein Teilbereich von Strukturgleichungsmodellen und wird zur Prüfung von Beziehungen zwischen Indikatorvariablen und hypothetischen Konstrukten bzw. latenten Variablen verwendet.¹² Im Gegensatz zur explorativen Faktorenanalyse, welches ein struktur-entdeckendes Verfahren ist, ist die KFA ein struktur-prüfendes Verfahren. Mit Hilfe der KFA wird daher eine a priori festgelegte Zuordnung von empirischen Indikatoren zu Faktoren (Konstrukten) überprüft, wobei die Zuordnung auf Basis von theoretischen oder zumindest sachlogisch fundierten Überlegen getroffen werden muss. Anders gesagt, wird ein bestimmtes Messmodell, welches die im Vorfeld einer Befragung getroffenen Überlegungen zur Abbildung eines Konstrukts anhand verschiedener Fragen bzw. Items, welche die Indikatorvariablen darstellen, enthält, empirisch überprüft (Backhaus et al. 2008A, S. 7-9). Latente Variablen müssen durch Messmodelle erfasst werden, weil sie auf der empirischen Ebene nicht direkt beobachtbar sind.

Entscheidend für die mögliche Anwendung der KFA ist zunächst, welche Richtung die Beziehung zwischen Indikatorvariablen und latenten Variablen besitzt. So kann die KFA nur bei reflexiven Messmodellen eingesetzt werden, die davon ausgehen, dass Veränderungen in den Messwerten der Indikatorvariablen durch die latente Variable kausal verursacht werden.¹³ Damit wird

¹²Außerdem können auf Basis der KFA empirische Messwerte für die latenten Variablen ermittelt werden.

¹³Entsprechend kann die KFA nicht bei formativen Messmodellen, die von Indikatorvariablen als Determinanten der latenten Variablen ausgehen, angewendet werden.

auch impliziert, dass Indikatorvariablen zur Messung einer latenten Variablen austauschbar sind. Formal kann die Beziehung allgemein wie folgt dargestellt werden (Backhaus et al. 2008A, S. 4-5):

$$x_i = \lambda_i \xi + \delta_i \quad (i = 1, \dots, n). \quad (3.1)$$

Damit entspricht der Wert eines Items x_i der Summe aus dem gewichteten (λ_i) Faktorwert (ξ) und einem Fehlerterm (δ_i).¹⁴ Die Gewichtungsgröße quantifiziert die Stärke des Zusammenhangs zwischen dem Faktor und der i-ten Indikatorvariablen und wird im Rahmen der KFA als Faktorladung beschrieben. Diese Größe wird auch als Korrelation zwischen einem Indikator und dem betrachteten Konstrukt interpretiert (Backhaus et al. 2008A, S. 5). Der Fehlerterm steht für die Fehlervarianz, die den durch die latente Variable unerklärte Varianzanteil eines Items bemisst. Die Formel (3.1) gilt unter der Annahme, dass die Werte zentriert oder z-standardisiert sind¹⁵, und enthält somit keine Konstante (Bühner 2006, S. 241).

Nach der Formulierung eines Modells muss im nächsten Schritt geprüft werden, ob die Modellparameter eindeutig bestimmt werden können. Ein Modell gilt als identifizierbar, wenn gilt:

$$t \leq \frac{1}{2} p(p+1) \quad (3.2)$$

mit t = Anzahl der zu schätzenden Modellparameter
und p = Anzahl der empirischen Indikatoren.

Aus p empirischen Indikatoren lassen sich $\frac{1}{2} p(p+1)$ empirische Varianzen und Kovarianzen errechnen, die zur Bestimmung der zu schätzenden Para-

¹⁴Das Modell entspricht dem Grundgedanken der klassischen Testtheorie: Beobachteter Wert = wahrer Wert + Messfehler (Bühner 2006, S. 241).

¹⁵z-Standardisierung: $\sum[(x_i - \bar{x})/\sigma]$. Eine Standardisierung der erhobenen Daten ist nicht erforderlich, um ein Modell zu spezifizieren oder dessen Parameter zu schätzen. Sie dient vielmehr zur anschaulichen Erläuterung der Rechenschritte bei der KFA.

meter in einem Modell bei der KFA verwendet werden. So muss die Anzahl der beobachteten Varianzen und Kovarianzen mindestens so groß sein wie die Anzahl der zu schätzenden Parameter, damit ein Modell identifizierbar ist. Weiterhin sollten auch noch genügend Informationen (sog. Freiheitsgrade ($d.f. > 0$)¹⁶) zur Bestimmung der Güte des Modells zur Verfügung stehen und daher sollte die Anzahl der empirischen Varianzen und Kovarianzen größer als die Anzahl der zu schätzenden Parameter sein (Backhaus et al. 2008A, S. 15). Somit sind drei Fälle zu unterscheiden:

1. Unteridentifizierte Modelle ($d.f. < 0$)

→ Es gibt mehrere Lösungen für die Parameterschätzung. Nur durch Parameterrestriktionen kann das Problem gelöst werden.

2. Gerade identifizierte Modelle ($d.f. = 0$)

→ Es gibt eine einzige Lösung für die Parameterschätzung, es kann allerdings kein Modelltest durchgeführt werden.

3. Überidentifizierte Modelle ($d.f. > 0$)

→ Es bestehen verschiedene Möglichkeiten die Modellparameter näherungsweise zu schätzen, ein Modelltest kann durchgeführt werden.

Im Falle eines überidentifizierten Modells werden die Modellparameter gesucht, durch die die Differenz bzw. die sog. Diskrepanz zwischen den empirischen Kovarianzen und den durch das Modell implizierten Kovarianzen minimiert wird (Bühner 2006, S. 242/243). Wie dies im Einzelnen geschieht wird später erläutern.

Nach der Prüfung, ob die Modellparameter eindeutig bestimmt werden können, und bevor die Parameter geschätzt werden, muss noch die Metrik bzw. Skala der latenten Variablen und Fehlervariablen festgelegt werden, um diese nicht beobachtbaren Größen nach ihrer Schätzung auch interpretieren zu

¹⁶ $d.f. = [\frac{1}{2} p (p + 1)] - t$. In Worten: Freiheitsgrade = Anzahl der empirischen Varianzen und Kovarianzen – Anzahl der zu schätzenden Parameter.

können. Da die Messwerte der Parameter mit Hilfe der Indikatorvariablen geschätzt werden, bieten sich die Indikatorvariablen als Referenz an. Diese Festlegung der Metrik kann auf zwei Arten geschehen:

1. Indem eine feste Zuweisung einer Indikatorvariablen zu einer latenten Variablen als Referenzvariable erfolgt bzw. indem die Faktorladung einer Indikatorvariablen als eins gesetzt wird. Damit ist die latente Variable bis auf den Messfehler mit der Indikatorvariablen identisch. Als Referenzvariable sollte deshalb der Indikator bestimmt werden, der die latente Variable am besten widerspiegelt oder eine hohe Reliabilität mit der latenten Variable aufweist. Analog können auch die Pfade der Fehlervariablen auf die Indikatorvariablen auf 1 fixiert werden.
2. Indem die Varianz einer latenten Variablen auf 1 gesetzt wird.¹⁷ Damit besteht der Vorteil, dass die freien Faktorladungen für alle Indikatorvariablen geschätzt werden, und außerdem, dass im Falle von Modellen mit mehreren latenten Variablen die Kovarianz dieser latenten Variablen der Korrelation entspricht.

Die beiden Möglichkeiten zur Bestimmung der Metrik der zu schätzenden Parameter können zu unterschiedlichen Ergebnissen führen, da die Anzahl der zu schätzenden Parameter bei gleicher Informationsmenge variiert. Führen beide Bestimmungsmethoden der Metrik zu gleichen oder sehr ähnlichen Parameterschätzungen kann von geeigneten Messungen der nicht beobachtbaren Variablen ausgegangen werden (Backhaus et al. 2008A, S. 16).

Nun wird mit Hilfe der KFA versucht, die empirischen Kovarianzen der Indikatorvariablen durch adäquate Schätzungen der Parameter des Modells möglichst genau abzubilden. Die KFA errechnet daher zunächst aus den erhobenen Daten die empirische Kovarianzmatrix (S) bzw. die empirische Korrelationsmatrix (R) der Indikatorvariablen. Die Korrelationsmatrix entspricht grundsätzlich der Kovarianzmatrix, wenn die als Berechnungsgrund-

¹⁷Diese Fixierung ist nicht geeignet für Multigruppenanalysen, bei denen die Gültigkeit eines Modells bei verschiedenen Gruppen überprüft wird (Bühner 2006, S. 244).

lage dienenden Werte standardisiert¹⁸ sind (Backhaus et al. 2008A, S. 17). Basierend auf dem Fundamentaltheorem der Faktorenanalyse¹⁹ nach Thurstone (1931) und unter Verwendung des aufgestellten Gesamtmodells²⁰ lässt sich die sog. modelltheoretische Kovarianz- bzw. Korrelationsmatrix (Σ) wie folgt darstellen:

$$\Sigma = \Lambda \Phi \Lambda^\top + \Theta \quad (3.3)$$

mit Λ = Matrix der Faktorladungen, Φ = Kovarianz- bzw. Korrelationsmatrix der Faktorwerte und Θ = Kovarianz- bzw. Korrelationsmatrix der Fehlerterme.

Die modelltheoretische Kovarianzmatrix ergibt sich in der beschriebenen Weise nur, wenn die beiden Voraussetzungen erfüllt sind, dass die Fehlerterme sowohl untereinander als auch mit den entsprechenden latenten Variablen nicht korrelieren (Bühner 2006, S. 246). Nachdem die empirische Kovarianzmatrix (S) berechnet ist, wird sie mit der modelltheoretischen Kovarianzmatrix (Σ) gleichgesetzt und so die Parameter geschätzt. Je geringer die Diskrepanz zwischen den empirischen und den modelltheoretischen Kovarianzen ist, desto verlässlicher sind die Parameterschätzungen und desto besser ist die Güte des Modells (Backhaus et al. 2008A, S. 17). Die Parameter des Modells werden wie bereits beschrieben nur im Fall von gerade identifizierten Modellen eindeutig bestimmt. Im Fall von überidentifizierten Modellen werden die Modellparameter iterativ mit verschiedenen Startwerten so geschätzt, dass die Diskrepanz minimiert wird.

Die Diskrepanz bzw. die gewichtete Differenz zwischen den beobachteten und modelltheoretischen Kovarianzen kann durch verschiedene Methoden ermittelt werden. Folgende, üblicherweise verwendeten Schätzmethoden stehen zur

¹⁸Daraus folgt: Mittelwert = 0, Varianz = 1.

¹⁹Fundamentaltheorem der Faktorenanalyse: $\Sigma = \Lambda \Phi \Lambda^\top$. Ein wichtiger Zwischenschritt bei der Herleitung des Theorems beruht auf der Tatsache, dass Faktorwerte sich als Lineartransformation aus den empirischen Daten ergeben und daher aus standardisierten Daten errechnete Faktorwerte auch standardisiert sind (Backhaus et al. 2008A, S. 6/7).

²⁰Basierend auf der in Formel 3.1 dargestellten allgemeinen Modellgleichung.

Verfügung (Backhaus et al. 2006, S. 362/363, Browne 1984):

- Maximum-Likelihood-Methode (ML):

$$F_{ML} = \log|\Sigma| + \text{tr}(S \cdot \Sigma^{-1}) - \log|S| - (p + q) \quad (3.4)$$

- Methode der ungewichteten kleinsten Quadrate (ULS):

$$F_{ULS} = \frac{1}{2} \cdot \text{tr}(S - \Sigma)^2 \quad (3.5)$$

- Methode der verallgemeinerten kleinsten Quadrate (GLS):

$$F_{GLS} = \frac{1}{2} \cdot \text{tr}[S^{-1} \cdot (S - \Sigma)]^2 \quad (3.6)$$

- Methode der skalenunabhängigen kleinsten Quadrate (SLS):

$$F_{SLS} = \frac{1}{2} \cdot \text{tr}[\text{diag}(S)^{-1} \cdot (S - \Sigma)]^2 \quad (3.7)$$

- Methode der asymptotisch verteilungsfreien Schätzer (ADF)²¹:

$$F_{ADF} = [\text{vec}(S) - \text{vec}(\Sigma(\pi))]^\top U^{-1} [\text{vec}(S) - \text{vec}(\Sigma(\pi))] \quad (3.8)$$

$$U = s_{ijkl} - s_{ij}s_{kl} \quad (3.9)$$

mit p = Anzahl der Indikatorvariablen, q = Anzahl der zu schätzenden Parameter, π = Parametervektor der Länge q , tr = Summe der Diagonalelemente (*trace*) einer quadratischen Matrix, diag = Diagonalelemente einer quadratischen Matrix, vec = Matricelemente werden als einspaltiger Vektor geschrieben, s_{ijkl} = multivariater Momente vierter Ordnung und s_{ij} = Moment zweiter Ordnung der Variablen x_i und x_j (s_{kl} analog).

Die verschiedenen Methoden bzw. Diskrepanzfunktionen müssen in Abhän-

²¹Hinweis zu Formel 3.9: $s_{ijkl} = \frac{\sum (x_i - \bar{x}_i)(x_j - \bar{x}_j)(x_k - \bar{x}_k)(x_l - \bar{x}_l)}{n}$ und $s_{ij} = \frac{\sum (x_i - \bar{x}_i)(x_j - \bar{x}_j)}{n}$ (Arbuckle 2007, S. 582/583).

gigkeit der Stichprobengröße²² und der Verteilungsannahme bezüglich der Indikatorvariablen eingesetzt werden. Außerdem ist relevant, ob die Schätzfunktion skaleninvariant²³ ist und ob Inferenzstatistiken zur Beurteilung von Modellen zu Verfügung stehen. Tabelle 3.1 verdeutlicht die Voraussetzungen der einzelnen Methoden (Backhaus et al. 2006, S. 363-365).

Tab. 3.1: Kriterien und Besonderheiten der Schätzverfahren im Rahmen der KFA

Voraussetzung	ML	ULS	GLS	SLS	ADF
Multinormalverteilung	ja	nein	ja	nein	nein
Stichprobengröße	>100	>100	>100	>100	>500
Skaleninvarianz	ja	nein	ja	ja	ja
Inferenzstatistiken (χ^2)	ja	nein	ja	nein	ja

Trotz zum Teil sehr unterschiedlicher Aussagen bezüglich der Anwendbarkeit der verschiedenen beschriebenen Schätzverfahren wird in der Regel die ML-Methode zur Schätzung der Parameter verwendet. Sowohl die ULS-Methode als auch die SLS-Methode kommen u.a. auf Grund der geringen Anzahl an Möglichkeiten zur Bestimmung der Modellgüte in der Praxis kaum zum Einsatz. Die ML-Methode ist der GLS-Methode vorzuziehen, da die ML-Methode zu einer höheren Genauigkeit bei der Schätzung der Parameter führt (Olsson et al. 2000). Eine Entscheidung zwischen der ML- und der ADF-Methode könnte am einfachsten anhand der Verteilungsannahme und der Skalierung der Daten getroffen werden. Die ML-Methode ist unter der Annahme einer Multinormalverteilung und intervallskalierten Daten²⁴ in jedem Fall anzuwenden (Schermerle-Engel et al. 2003, S. 49). Weiterhin ist nach

²²Die minimale Stichprobengröße wird im Bezug auf die Anwendbarkeit und die Genauigkeit der Ergebnisse in der Literatur z.T. sehr unterschiedlich beurteilt (Boomsma 1982, Nevitt/Hancock 2001). Besonders hervorzuheben ist, dass die ADF-Methode nur bei sehr großen Stichproben (Nevitt/Hancock 2001: Stichprobengröße > 5.000) anzuwenden ist.

²³Skaleninvariant bedeutet, dass die Skalierung keine Auswirkungen auf die Ergebnisse der Analyse hat. Im Zusammenhang mit der KFA zielt Skaleninvarianz vor allem darauf ab, dass sowohl die Verwendung der Kovarianzmatrix als auch die Verwendung der Korrelationsmatrix zu identischen Ergebnissen führt.

²⁴Zumindest muss eine Intervallskalierung der Daten unterstellt werden können (Bühner 2006, S. 251)

Olsson et al. (2000) selbst bei Verletzung der Multinormalverteilung und einer Stichprobe mit 2.000 Beobachtungen die ML-Methode im Bezug auf die Genauigkeit der Parameter und die Modellgüte der ADF-Methode vorzuziehen. Die ML-Methode ist somit relativ robust gegenüber Verletzungen der Multinormalverteilungsannahme und es stehen auch in einem solchen Fall robuste Verfahren zur Beurteilung der Modellgüte zur Verfügung. Hingegen sollte die ADF-Methode zur Parameterschätzung bei Verletzung der multivariaten Normalverteilung, ordinalskalierten Daten und extrem großen Stichproben durchgeführt werden (Bühner 2006, S. 250-252). Da eine Entscheidung zwischen der ML- und der ADF-Methode im Vorfeld schwierig ist, muss eine Abwägung vor allem anhand der multivariaten Verteilung, aber auch unter Berücksichtigung der Skalierung der Daten der Stichprobe und der Stichprobengröße erfolgen.

Vor der Untersuchung der multivariaten Verteilung sollte die univariate Verteilung geprüft werden, da eine univariate Normalverteilung eine notwendige Bedingung für das Vorliegen von multivariater Normalverteilung ist. Üblicher Weise wird bei der KFA das Vorliegen von univariate Normalverteilung mit Hilfe der Schiefe und der Kurtosis (Exzess) bestimmt (DeCarlo 1997, S. 296/297). Die Schiefe und die Kurtosis werden in folgender Weise berechnet (Hartung et al. 2005, S. 47ff.):

$$Schiefe = \frac{\sum \left(\frac{x_i - \bar{x}}{\sigma} \right)^3}{n} \quad (3.10)$$

$$Kurtosis = \frac{\sum \left(\frac{x_i - \bar{x}}{\sigma} \right)^4}{n} \quad (3.11)$$

mit x_i = i-te Beobachtung der Variable, \bar{x} = Mittelwert der Variable, σ = Standardabweichung der Variable und n = Anzahl der Beobachtungen.

Nach Curran et al. (1996) sollte man bei einer Schiefe größer als zwei und einer Kurtosis größer als sieben keine Schätzmethode mit multivariater Normalverteilungsannahme verwenden. Da eine univariate Normalverteilung nur eine notwendige Bedingung ist, muss u.U. auch die multivariate Verteilung

überprüft werden. Ein Maß zur Einschätzung der multivariaten Verteilung ist die multivariaten Kurtosis nach Mardia (1970, 1974), die sich aus der folgender Formel ergibt:

$$\text{multivariate Kurtosis} = \frac{1}{n} \sum_{i=1}^n \left[(x_i - \bar{x})^\top \hat{S}^{-1} (x_i - \bar{x}) \right]^2 - \frac{p(p+2)(n-1)}{n+1} \quad (3.12)$$

mit x_i = i-te Beobachtung der p Variablen, \bar{x} = Vektor der jeweiligen Mittelwerte, \hat{S} = unverzerrte Schätzung der Kovarianzmatrix der Grundgesamtheit und n = Anzahl der Beobachtungen.

Unter der Annahme einer normalverteilten Grundgesamtheit ist der Mittelwert der Kurtosis nach Mardia null und der Standardfehler gleich $[8p(p+2)/n]^{1/2}$. Ist das „Critical Ratio“, das als z-Wert des sog. Mardia-Tests interpretiert wird und das aus dem Quotient der multivariaten Kurtosis und dem Standardfehler gebildet wird, größer als 1,96 kann man auf einem 95%-Niveau bei einem zweiseitigen Test nicht von multivariater Normalverteilung ausgehen.

Nachdem die Parameter mit Hilfe der ML- oder ADF-Methode geschätzt wurden, kann die Güte des Modells durch inferenzstatistische und deskriptive Gütekriterien beurteilt werden. Als inferenzstatistisches Gütekriterium wird der χ^2 -Test verwendet. Beim χ^2 -Test wird die Nullhypothese $H_0: S = \Sigma$ gegen die Alternativhypothese $H_1: S \neq \Sigma$ geprüft. Die χ^2 -Prüfgröße wird mit Hilfe der verwendeten Diskrepanzfunktion in folgender Weise berechnet (Backhaus et al. 2008A, S. 27):

$$\chi^2 = (N - 1) F_j \quad (3.13)$$

mit F_j = Verwendete Diskrepanzfunktion j und N = Stichprobenumfang.

Ist die dazugehörige Wahrscheinlichkeit p (*probability level*) größer als 0,05

wird die Nullhypothese angenommen und das theoretische Modell kann durch die empirischen Daten bestätigt werden. Daraus kann allerdings nicht geschlossen werden, dass andere Modelle nicht mindestens genauso gut zu den Daten passen. Der χ^2 -Test ist außerdem nur zur Bestimmung der Güte eines Modells im Rahmen der KFA geeignet, falls die Voraussetzungen von multinormalverteilten Indikatorvariablen und einer adäquaten Stichprobengröße²⁵ erfüllt sind. Weiterhin kann die Prüfgröße des Tests durch die Komplexität bzw. die Anzahl der Parameter des Modells²⁶ beeinflusst sein (Schermelleh-Engel et al. 2003, S. 31-33). Falls die Annahme einer Multinormalverteilung verletzt sein sollte und damit die Gültigkeit des χ^2 -Test in Frage gestellt ist, besteht die Möglichkeit eine korrigierte Übertretungswahrscheinlichkeit (p) mit der Bollen-Stine-Bootstrap-Methode²⁷ zu errechnen (Bühner 2006, S. 251).

Auf Grund einer eingeschränkten Gültigkeit des χ^2 -Tests und dessen korrigiertem Wert besonders bei großen Stichproben sollten auch noch deskriptive Gütekriterien (sog. Fit-Indizes) zur Beurteilung eines Modells verwendet werden. Grundsätzlich gibt es zwei Typen von deskriptiven Gütekriterien: Absolute Fit-Indizes und komparative Fit-Indizes. Bei absoluten Fit-Indizes wird ein Vergleich mit einem sog. „saturated“ Modell vorgenommen, das die Stichprobenkovarianz genau nachbildet. Hingegen wird bei komparativen Fit-Indizes²⁸ ein Vergleich mit dem sog. „Independence“-Modell, bei dem nur die Varianzen der empirischen Indikatorvariablen geschätzt werden, weil alle anderen Parameter auf null gesetzt werden, vorgenommen (Bühner 2006, S. 254/255). Hu und Bentler (1998) empfehlen sowohl bei Verwendung der ML-

²⁵Mit der Größe der Stichprobe unter der Bedingung einer konstanten Zahl an Freiheitsgraden steigt der χ^2 -Wert und plausible Modelle werden u.U. abgelehnt. Diese Problematik kann auch nicht durch die deskriptive Beurteilung der Modellgüte mit Hilfe des Verhältnisses von χ^2 -Prüfgröße und Anzahl der Freiheitsgrade gelöst werden.

²⁶Komplexe Modelle haben tendenziell kleiner χ^2 -Werte als einfachere Modelle.

²⁷Für weitergehende Informationen zur Bollen-Stine-Bootstrap Methode siehe Bollen und Stine (1993)

²⁸Bei den komparativen Fit-Indices gibt es weiterhin eine Einteilung in drei verschiedene Typenklassen, die jeweils eine unterschiedlich hohe Anzahl an Parametern bei der Berechnung des jeweiligen Indizes berücksichtigen. Typ-III-Indizes berücksichtigen dabei die meisten Parameter.

als auch der ADF-Schätzmethode in erster Linie das folgende deskriptive Gütekriterium:

- Standardized-Root-Mean-Residual (SRMR):

$$SRMR = \sqrt{2 \sum_{i=1}^p \sum_{j=1}^i \frac{r_{ij}^2}{p(p+1)}} \quad (3.14)$$

mit r_{ij} = Korrelation zwischen Item i und j , p = Anzahl der Items.

Hu und Bentler (1999, S. 27) sprechen im Zusammenhang mit dem SRMR von einem relativ guten Fit zwischen dem angenommenen Modell und den empirischen Daten bei SRMR-Werten kleiner 0,08. Außerdem empfehlen Hu und Bentler (1998) noch die folgende komparativen Fit-Indizes zur ergänzenden Beurteilung der Modellgüte:

- Comparative-Fit-Index (CFI, Typ III):

$$CFI = 1 - \frac{\max[(\chi_T^2 - d.f.T), 0]}{\max[(\chi_T^2 - d.f.T), (\chi_I^2 - d.f.I), 0]}. \quad (3.15)$$

- Tucker-Lewis-Index (TLI, Typ II):

$$TLI = \frac{(\chi_I^2/d.f.I) - (\chi_T^2/d.f.T)}{(\chi_I^2/d.f.I) - 1} \quad (3.16)$$

mit T = geprüftes Modell, I = Independence-Modell und $d.f.$ = Freiheitsgrade.

Der CFI nimmt Werte zwischen null und eins an und bei Werten ab 0,95 bzw. 0,97 spricht man von einem akzeptablen bzw. guten Modell-Fit. Weiterhin ist der CFI relativ robust gegenüber der Stichprobengröße und Verteilungsverletzungen. Beim TLI gelten dieselben Werte bezüglich des Modell-Fits und auch der TLI ist relativ robust gegenüber der Stichprobengröße. Im Gegensatz zum CFI ist der TLI nicht normiert (Schermelleh-Engel et al. 2003, S. 40-42, Backhaus et al. 2008A, S. 32).

Nachdem die Güte eines Modells mit Hilfe des χ^2 -Tests und den beschriebenen Indices beurteilt wurde, kann bei unzureichender Modellgüte auf Basis der Berechnungen der KFA und von weiteren sachlogischen Überlegungen eine Modifikation eines Messmodells vorgenommen werden. Zwar verliert damit die KFA den reinen Charakter eines struktur-prüfenden Verfahrens, jedoch besteht im Rahmen der hier durchgeführten KFAs nicht nur die Zielsetzung, ob die Abbildung von bestimmten Konstrukten anhand der zugeordneten Items gelungen ist. Vielmehr dient die KFA als Basis zur anschließenden Untersuchung von Zusammenhängen zwischen durch die Daten bestätigten Konstrukten. Um dabei möglichst viele Aspekte bzw. durch die Daten messbare Konstrukte verwenden zu können, wird hier ggf. versucht durch die Reduktion oder Erweiterung der zunächst gewählten Messmodelle möglichst viele adäquat gemessene Konstrukte identifizieren zu können.

3.2 Strukturgleichungsmodelle

Mit Hilfe von Strukturgleichungsmodellen (SGM) können kausale Abhängigkeiten²⁹ zwischen Merkmalen bzw. Variablen analysiert werden, wobei im Vorfeld der statistischen Anwendung auf Basis von sachlogisch oder theoretisch begründeten Überlegungen Annahmen über die Art und Weise der Beziehungen getroffen werden. Dieses sog. theoretisch fundierte Hypothesensystem bzw. Strukturmodell wird dann dahingehend geprüft, ob die theoretisch aufgestellten Beziehungen durch die erhobenen Daten bestätigt werden können. Im Gegensatz zu anderen statistischen Methoden der Kausalanalyse können mit Hilfe von SGM auch lineare Beziehungen zwischen endogenen (abhängigen) und exogenen (unabhängigen) latenten Variablen untersucht werden. Wie bereits bei der KFA erwähnt können latente Variablen bzw. hypothetischen Konstrukte nicht direkt beobachtet werden und müssen daher

²⁹Eine kausale oder direkte Abhängigkeit zwischen einer Variablen X und einer Variablen Y besteht, wenn Variationen der Variablen Y durch Variationen der Variablen X hervorgerufen werden bzw. eine klare Zuordnung von Ursache und Wirkung möglich ist (Backhaus et al. 2008B, S. 7).

mit Hilfe von adäquaten Indikatoren³⁰ indirekt durch Messmodelle quantifiziert werden (Backhaus et al. 2008B, S. 3/4). So werden in einem SGM die aufgestellten Beziehungen zwischen hypothetischen Konstrukten, die durch Messmodelle erfasst werden, im Bezug auf ihre empirische Gültigkeit überprüft.

Die auf Basis von sachlogischen Überlegungen getroffenen Annahmen bezüglich der Abhängigkeiten zwischen endogenen und exogenen latenten Variablen werden in einem sog. Strukturmodell wie folgt operationalisiert (Backhaus et al. 2008B, S. 10):

$$\eta = \Gamma\xi + \Psi \quad (3.17)$$

mit η = Matrix der latenten endogenen Variablen, ξ = Matrix der latenten exogenen Variablen, Γ = Matrix der partiellen Regressionskoeffizienten und Ψ = Matrix der Residualvariablen für die endogenen latenten Variablen η_i .

Der Formel liegen die Annahmen zugrunde, dass nur eine endogene latente Variable untersucht wird und dass die empirischen Daten standardisiert sind. Weiterhin müssen auf Grund der nicht direkten Beobachtbarkeit der latenten Variablen Messmodelle³¹ mit einem oder mehreren Indikatoren aufgestellt werden. Im Rahmen von SGM unterscheidet man zwischen Messmodellen für endogene latente Variablen (η) und Messmodellen für exogene latente Variablen (ξ). Allgemein lassen sich die Messmodelle demnach in folgender Weise formulieren (Backhaus et al. 2008B, S. 11/12):

³⁰In wie weit Indikatoren zur Abbildung eines hypothetischen Konstrukts geeignet sind, kann mit Hilfe der KFA überprüft werden.

³¹Messmodelle wurden bereits in Abschnitt 3.1 erläutert und werden auch im Rahmen von SGM analog zu Formel 3.1 formuliert. Es gelten die in Abschnitt 3.1 beschriebenen Annahmen und Bedingungen, beispielsweise bezüglich der Metrik, auch im Rahmen von SGM.

$$X = \Lambda_x \xi + \delta \quad (3.18)$$

$$Y = \Lambda_y \eta + \epsilon \quad (3.19)$$

mit X = Matrix der Indikatorvariablen der latenten exogenen/unabhängige Variable, Y = Matrix der Indikatorvariablen der latenten endogenen/abhängige Variablen, Λ_x sowie Λ_y = Matrizen der jeweiligen partiellen Regressionskoeffizienten, ξ = Matrix der latenten exogenen Variablen, η = Matrix der latenten endogenen Variablen, δ = Matrix der Residualvariablen für die Indikatorvariablen x_i und ϵ = Matrix der Residualvariablen für die Indikatorvariablen y_i .

Nachdem die kausalen Abhängigkeiten zwischen den latenten Variablen operationalisiert wurden und damit Annahmen bzw. Hypothesen bezüglich der Richtung der Abhängigkeiten auf Basis von sachlogischen Überlegungen getroffen wurden, können grundsätzlich die Parameter aus den empirischen Daten geschätzt und damit die aufgestellten Hypothesen empirisch überprüft werden. Dabei wird von den Annahmen untereinander unkorrelierter Residual- bzw. Messfehlervariablen (d.h. keine Korrelationen zwischen ζ , δ und ϵ) und nicht korrelierenden Messfehlervariablen und hypothetischen Konstrukten (d.h. ζ korreliert nicht mit ξ , δ nicht mit ξ und ϵ nicht mit η) ausgegangen (Backhaus et al. 2008B, S. 19). Wie bei der KFA sind auch bei SGM die aufgestellten Gleichungssysteme identifizierbar bzw. lösbar, wenn die Anzahl der empirischen gegebenen Parameter mindestens genauso groß ist wie die Anzahl der zu schätzenden Parameter (Backhaus et al. 2008B, S. 21/22).³² Wenn die Anzahl der empirischen Parameter gleich der Anzahl der zu schätzenden Parameter ist, handelt es sich um ein gerade identifiziertes Modell und alle zu schätzenden Parameter können eindeutig bestimmt werden. Übersteigt die Anzahl der empirischen Parameter die Anzahl der zu schätzenden Parameter spricht man von überidentifizierten Modellen und

³²Basierend auf Formel 3.2 kann diese Bedingung bei SGM in folgender Weise formalisiert werden: $t \leq \frac{1}{2} (p + q) (p + q + 1)$, mit t = Anzahl der zu schätzenden Modellparameter, p = Anzahl der empirischen endogenen Indikatoren und q = Anzahl der empirischen exogenen Indikatoren.

es bestehen mehrere Möglichkeiten die Modellparameter zu schätzen. Eine weitere Bedingung für die Identifizierbarkeit eines Modells ist, dass die zu schätzenden Gleichungen linear unabhängig bzw. dass die zur Schätzung notwendigen Matrizen invertierbar sind (Backhaus et al. 2008B, S. 22).³³

Im Rahmen dieser Arbeit werden Strukturgleichungsmodelle durch eine simultane Schätzung aller Modellparameter gelöst. Bei diesem sog. faktor- oder kovarianzanalytischen Ansatz werden die Kovarianzen bzw. Korrelationen der Indikatorvariablen berechnet und dann zur Bestimmung der Beziehungen zwischen den Indikatorvariablen und latenten Variablen sowie zwischen den latenten abhängigen und unabhängigen Variablen verwendet.³⁴ Damit werden die Beziehungen mit festgelegter Richtung zwischen den latenten Variablen nicht direkt auf Basis der erhobenen Daten, sondern auf Basis der aus den empirischen Daten berechneten Kovarianzen bzw. Korrelationen geschätzt und das Strukturmodell insgesamt untersucht. Eine weitere Vorgehensweise zur Schätzung eines Strukturgleichungsmodells ist der sog. varianzanalytische Ansatz, bei dem zunächst die Faktorladungen der Messmodelle mit der KFA geschätzt und dann auf Basis derer die Faktorwerte bzw. Beobachtungswerte der latenten Variablen berechnet werden. Anschließend wird mit Hilfe der berechneten Faktorwerte eine Regressionsanalyse durchgeführt und durch die geschätzten Regressionskoeffizienten können Aussagen über die Beziehungen zwischen den abhängigen und unabhängigen latenten Variablen getroffen werden (Backhaus et al. 2008B, S. 4/5). In dieser Arbeit wird der sog. kovarianzanalytische Ansatz gewählt, da dieser Ansatz in dem im Rahmen der KFA verwendeten Statistikprogramm AMOS implementiert ist und keine sehr umfangreichen Berechnungen der Faktorwerte notwendig sind.

³³Damit ein Strukturgleichungsmodell lösbar ist, muss insbesondere die empirische Korrelationsmatrix positiv definit bzw. invertierbar sein.

³⁴Da durch die Kovarianz die Stärke des Zusammenhangs zwischen zwei Variablen grundsätzlich nicht bestimmt werden kann, wird die Kovarianz durch die Standardabweichungen der Variablen dividiert und damit die Korrelation zwischen zwei Variablen berechnet. Mit Hilfe der Korrelation kann dann die Stärke des Zusammenhangs beziffert werden. Bei standardisierten Daten ist wie bereits erwähnt die Varianz-Kovarianz-Matrix gleich der Korrelationsmatrix (Backhaus et al. 2008B, S. 7).

Analog zu der Verfahrensweise bei der KFA kann auch bei den Strukturgleichungsmodellen die modelltheoretische Korrelationsmatrix (Σ) mit Hilfe der aufgestellten Modellgleichungen und dem Fundamentaltheorems der Faktorenanalyse formuliert werden (siehe Formel 3.3). Die durch das Modell implizierte Korrelationsmatrix (Σ) wird dann mit der empirischen Korrelationsmatrix (S) gleichgesetzt und damit die zu schätzenden Modellparameter bestimmt. Dabei ist wieder die Zielsetzung die Differenz zwischen den modelltheoretischen und den empirischen Korrelationen zu minimieren. Bei gerade identifizierten Strukturgleichungsmodellen gibt es für die einzelnen Modellparameter genau eine Lösung. Hingegen müssen bei überidentifizierten SGM die Parameter des Modells iterativ geschätzt werden, da es keine eindeutige Lösung für die zu schätzenden Parameter gibt. Die Differenzen zwischen den modelltheoretischen und empirischen Korrelationen können durch unterschiedliche Diskrepanzfunktionen abgebildet werden, wobei die Wahl der Diskrepanzfunktion bzw. Schätzmethode in Abhängigkeit der Verteilung und der Skalierung der Daten sowie der Stichprobengröße getroffen werden muss. Zur Auswahl stehen im Rahmen von SGM genau wie bei der KFA die Maximum-Likelihood-Schätzmethode (ML), die Methode der ungewichteten kleinsten Quadrate (ULS), die Methode der verallgemeinerten kleinsten Quadrate (GLS), die Methode der skalenunabhängigen kleinsten Quadrate (SLS) und die Methode der asymptotisch verteilungsfreien Schätzer (ADF; Backhaus et al. 2006, S. 368-376). In Abschnitt 3.1 sind die entsprechenden Formalisierungen dargestellt (siehe Formel 3.4 - 3.9). Die Voraussetzungen und Empfehlungen bezüglich der Anwendbarkeit unterscheiden sich auch nicht, weshalb auf eine erneute Erläuterung an dieser Stelle verzichtet wird.

Nachdem in der beschriebenen Weise und bei Verwendung einer adäquaten Schätzmethode alle Parameter des Strukturgleichungsmodells bestimmt werden können³⁵, können nun die Schätzergebnisse beurteilt werden. In einem ersten Schritt wird die Zuverlässigkeit der Parameterschätzungen überprüft,

³⁵Nur im Fall eines unteridentifizierten SGM können die Parameter nicht geschätzt werden. Allerdings wird im Rahmen der Modellformulierung bereits darauf geachtet, dass die Identifizierbarkeit des Modells gewährleistet ist, und damit die Möglichkeit eines nicht lösbaren Modells bereits im Vorfeld der Berechnungen ausgeschlossen.

d.h. in wie weit kongruieren unabhängige, aber vergleichbare Messungen, und ob Fehlspezifikationen im Modell vorliegen. Mit den Standardfehlern (S.E.) der Schätzung, die die Streuung bei den jeweiligen Parameterschätzungen angeben, wird ein Hinweis auf die Zuverlässigkeit der Schätzungen gegeben. Ein großer Standardfehler bedeutet, dass ein geschätzter Modellparameter bei verschiedenen Stichproben aus einer Grundgesamtheit stark variiert und daher die Schätzung des Parameters nicht zuverlässig ist. Die Zuverlässigkeit der Messungen der Indikatorvariablen und der latenten endogenen Variablen kann durch die sog. Reliabilität³⁶ beurteilt werden. Normalerweise nimmt der Reliabilitätskoeffizient Werte zwischen null und eins an, wobei höhere Werte für eine höhere Zuverlässigkeit der Messungen sprechen. Besitzt ein Reliabilitätskoeffizient allerdings Werte über eins kann von Fehlspezifikationen im Modell ausgegangen werden und diese Information dann als Anhaltspunkt zur Modifikation eines Modells genutzt werden. Die Reliabilität wird in dieser Arbeit in Abhängigkeit der gewählten Statistiksoftware (AMOS) mit Hilfe der quadrierten multiplen Regressionskoeffizienten gemessen und dient auch zur Beurteilung der Stärke der kausalen Beziehungen in den Strukturgleichungen. Bei Messmodellen wird die Stärke der Beziehung zwischen einer Indikatorvariablen und der dahinterstehenden latenten Variable nach der folgenden Formel berechnet (Backhaus et al. 2008B, S. 27/28):

$$rel(x_i) = \frac{\lambda_{ij}^2 \phi_{jj}}{\lambda_{ij}^2 \phi_{ij} + \theta_{ii}} \quad (3.20)$$

mit λ_{ij} = geschätzter Regressionskoeffizient, ϕ_{jj} = Varianz der latenten Variable und θ_{ii} = geschätzte Varianz des entsprechenden Messfehlers.

Wird die latente Variable eines Messmodells standardisiert, ergibt sich die Indikatorreliabilität aus dem Quadrat der jeweiligen Faktorladung ($rel(x_i) = \lambda_{ij}^2$). Die Indikatorreliabilität zeigt an, wie viel Prozent der Varianz einer Messvariablen durch das dahinterstehende hypothetische Konstrukt in dem Modell erklärt wird. Der Umgang mit den Werten des Reliabilitätskoeffizi-

³⁶Allgemeine Formel der Reliabilität: *Reliabilitätskoeffizient* = $1 - \frac{\text{Fehlervarianz}}{\text{Gesamtvarianz}}$.

enten bzw. Grenzwerte für die Eignung einer Indikatorvariablen werden im Datenteil erörtert werden.

Neben den bisherigen Methoden zur Beurteilung der Zuverlässigkeit von Messungen müssen auch die Korrelationen zwischen den Parameterschätzungen untersucht werden. Denn bestehen sehr hohe Korrelationen zwischen zwei Parametern, ist davon auszugehen, dass die Parameter deckungsgleiche Aspekte messen, und damit ein Parameter abkömmlich ist. Backhaus et al. (2008B, S. 27/28) gehen bei Korrelationen absolut über 0,9 von redundanten Parametern aus. Außerdem können hohe Korrelationen zwischen exogenen latenten Variablen ein Indiz für Multikollinearität sein. In der Literatur wird die Problematik von Multikollinearität bei SGM oft vernachlässigt, da keine Einigkeit darüber besteht, ob Multikollinearität überhaupt eine Problematik bei SGM darstellt (Grewal et al. 2004, S. 519). Ein Grund für die Uneinigkeit ist, dass hoch korrelierte Variablen als Indikator für ein allgemein zugrundeliegendes Konstrukt angesehen werden können und durch eine adäquate Modellformulierung, d.h. Zusammenfassung oder Entfernung von Variablen, Multikollinearitätsprobleme im Bezug auf die latente endogene Variable vermieden werden können (Maruyama 1998, S. 60 ff.). Weiterhin wird die Problematik von Multikollinearität in der Literatur z.T. vernachlässigt, da der Umgang mit Multikollinearität nicht eindeutig geklärt ist, die Gründe für Multikollinearität unklar sind und damit eine Vermeidung schwierig ist (Grewal et al. 2004, S. 520). Nach Grewal et al. (2004, S. 528) kann die Exaktheit der SGM-Schätzergebnisse durch Multikollinearität beeinflusst werden. Andererseits zeigen sie aber auch, dass bei einer hohen Reliabilität bzw. einem hohen Erklärungsgehalt und mit Hilfe einer großen Stichprobe die negativen Effekte von Multikollinearität bei SGM abgemildert werden und dass durch die Berücksichtigung von Messfehler Strukturgleichungsmodelle weniger anfällig gegenüber Multikollinearität sind als Regressionen. Im Rahmen dieser Arbeit wird auf Grund dieser Erkenntnisse vor allem anhand der Zuverlässigkeit der Messungen sowie den Korrelationen in einem Modell und unter Berücksichtigung der Sachlogik entschieden, ob durch Multikollinearität verursachte Schätzprobleme zu erwarten sind. Sollten Indizien für Schätzprobleme be-

obachtet werden können, werden entsprechenden Modifikationsversuche der SGM unternommen und untersucht.

Die Güte eines Strukturmodells in seiner Gesamtheit wird üblicherweise mit den gleichen Methoden wie die Güte eines Messmodells im Rahmen der KFA (siehe Formel 3.13 bis 3.16) beurteilt, weshalb hier keine erneute Auseinandersetzung erfolgt. Neben der Beurteilung der Gesamtstruktur eines Modells ist auch die Beurteilung von Teilstrukturen sinnvoll, da damit im Fall einer nicht mindestens akzeptablen Modellgüte auch Anhaltspunkte für die Verbesserung eines Modells geliefert werden. Aus der Differenz der empirischen und der modelltheoretischen Korrelationsmatrix ergibt sich die Matrix der im Modell nicht erklärten Residuen.³⁷ Da diese Residuen nicht standardisiert sind, führen Veränderungen der Skalierung über die Varianzen und Kovarianzen auch zu Veränderungen der Residuen. Daher werden die unstandardisierten Residuen durch ihre geschätzte Standardabweichung dividiert und man erhält die standardisierten Residuen bzw. residuellen Kovarianzen. Je niedriger der Residualwert einer Variablen ist, desto geringer ist der Kovarianz- bzw. Korrelationsanteil der Variablen, der nicht durch das Modell erklärt werden kann. Backhaus et al. (2008B, S. 31/32) nennen als Indiz für ein gutes Modell Residualwerte von höchstens 0,1. Auch das sog. „Critical Ratio“ (C.R.) der geschätzten Parameter, das sich aus der Division eines geschätzten unstandardisierten Parameterwerts und dessen Standardfehler der Schätzung ergibt, kann zur Beurteilung von Teilstrukturen verwendet werden. Unter der Annahme multinormalverteilter empirischer Daten kann mit dem C.R. als Prüfgröße durch einen t-Test geprüft werden, ob die geschätzten Werte nicht signifikant von Null verschieden sind (H_0). Ist ein C.R. absolut größer als 1,96 kann die Nullhypothese auf einem 95%-Niveau verworfen werden und es kann davon ausgegangen werden, dass der entsprechende Parameter eine entscheidende Aufgabe im Strukturgleichungsmodell hat (Backhaus et al. 2008B, S. 32).

³⁷In AMOS wird diese Matrix unter dem Namen „(Standardized) Residual Covariances“ geführt.

4 Datenanalyse

4.1 Beobachtbarkeit der Konstrukte

Im ersten Teil der Datenanalyse liegt der Fokus auf der Beurteilung, ob anhand des verwendeten Fragebogens die Beobachtung der gewählten Konstrukte gelungen ist bzw. durch welche Items die Konstrukte am besten abgebildet werden können. Diese Analyse wird mit Hilfe der KFA durchgeführt und bildet den Ausgangspunkt für die anschließende Analyse der Konstrukte. Eine adäquate Schätzmethode im Rahmen der KFA muss u.a. in Abhängigkeit der Verteilung ausgewählt werden. Auf Grund dessen wird zunächst die univariate Verteilung anhand der Schiefe und der Kurtosis (Exzess) und dann ggf. die multivariate Verteilung anhand des Mardia-Tests beurteilt. Diese Untersuchung sowie die Prüfung der Modellgüte basiert auf der in Tabelle 2.5 dargestellten Zuordnung von Indikatorvariablen und Konstrukten.

Wie bereits erwähnt ist die univariate Normalverteilung eine notwendige Bedingung für das Vorliegen von multivariater Normalverteilung. Da die univariaten Schiefe bei allen betrachteten Items der einzelnen Konstrukte nicht größer als zwei und die Kurtosis nicht größer als sieben ist und somit innerhalb der von West et al. (1996) geforderten Grenzen zur Annahme einer Normalverteilung liegen, ist die notwendige Bedingung für eine multivariate Normalverteilung erfüllt. Da es sich aber nur um eine notwendige Bedingung handelt, muss auf Grund der Ergebnisse der univariaten Verteilungsanalyse auch die multivariate Verteilung betrachtet werden. Die Ergebnisse der multivariaten Verteilungsanalyse sind in Tabelle 4.1 veranschaulicht. Im Gegensatz zur univariaten Verteilung ist die multivariate Kurtosis in einigen Fällen deutlich erhöht und die dazugehörigen „Critical Ratios“ (C.R.), die als z-Werte

des Mardia-Tests interpretiert werden können, liegen bei allen Konstrukten deutlich über dem Bereich zur Annahme einer multivariate Normalverteilung.³⁸ Die ML-Schätzmethode ist zwar wie bereits erwähnt robust gegenüber Verletzungen der multivariaten Normalverteilungsannahme, allerdings wird im Folgenden auf Grund einer nur schwer annehmbaren multivariaten Normalverteilung und der weiterhin schwierigen Entscheidung zugunsten einer Schätzmethode sowohl die ML- als auch die ADF-Schätzmethode angewendet.³⁹ Bei Verwendung der ML-Schätzmethode muss bei Verletzung der Normalverteilungsannahme in jedem Fall eine Bollen-Stine-Bootstrap-Korrektur des χ^2 -Tests zur adäquaten Beurteilung der Modellgüte verwendet werden.

Tab. 4.1: Prüfung der multivariaten Verteilung

Konstrukt	Multivariater Kurtosis	Mardia-Test (C.R.)
Fachkompetenz	8,302	108,458
Unterrichtsführung	22,793	145,296
Diagnostik	1,234	16,126
Sozialklima	4,593	47,432
Veranschaulichung	6,737	57,617
Interessantheit	3,726	38,479
Feedback	5,917	61,116
Materialien	8,392	86,676
Schüleraktivierung	-1,184	-12,225
Schülerautonomie	15,936	81,040
Mindestanforderungen	7,204	94,109
Leistungserwartungen	3,394	44,342
Binnendifferenzierung	6,254	81,700
Zeitmanagement	6,244	53,399
Selbstengagement	3,120	32,226
Leistungsfähigkeit	3,517	36,324
Interesse	4,917	50,781
Schulbeurteilung	4,537	59,268
Lernzuwachs	4,805	49,624

³⁸Ist das C.R. größer als 1,96 kann nicht von Multinormalverteilung ausgegangen werden. Dies wurde bereits in Abschnitt 3.1 erläutert.

³⁹Die uni- und multivariate Verteilung wird nur bei den zunächst angenommenen Messmodellen geprüft, weil anschließend beide möglichen Schätzmethode verwendet werden.

Nach der Prüfung der Verteilung und der daraus folgenden Wahl der Schätzmethoden wird die Güte der gewählten Modelle untersucht. Grundsätzlich lässt sich bei Betrachtungen der Ergebnisse im Rahmen der KFA feststellen, dass sowohl bei der Festlegung der Metrik durch die feste Zuweisung der Faktorladung einer Indikatorvariablen auf eins als auch durch die Zuweisung der Varianz einer latenten Variable auf eins die Ergebnisse im Bezug auf die Modellgüte und die Faktorladungen bei diesem Datensatz identisch sind. Damit zeigt sich in einem ersten Schritt, dass es sich bei den Modellen um grundsätzlich geeignete Messungen handelt. Außerdem hat die Wahl der Indikatorvariablen bei der die Faktorladung auf eins festgelegt wird grundsätzlich keinen Einfluss auf die Schätzungen. Trotzdem wird bei der Wahl der Metrik nach dieser Methode immer der Indikatorvariablen eine feste Faktorladung zugewiesen, die nach Wagner (2008) die höchste korrigierte Korrelation mit der Skala⁴⁰ aus den restlichen Indikatorvariablen des entsprechenden Konstrukts hat. Im Weiteren wird bei der KFA auf Grund dieser Erkenntnisse die Metrik nur noch thematisiert, falls abweichenden Ergebnissen auftreten sollten.

Misst man die Güte der aufgestellten Modelle durch die korrigierte Wahrscheinlichkeit (p-level) des χ^2 -Tests, können die theoretischen Modelle nicht durch die empirischen Daten bestätigt werden, da alle p-Werte deutlich unter fünf Prozent liegt. Einerseits kann mit Hilfe der Bollen-Stine-Bootstrap-Korrektur des χ^2 -Wertes die Modellgüte bei Verletzung der Multinormalverteilung und Verwendung der ML-Schätzmethode beurteilt werden, andererseits besteht weiterhin die Problematik, dass bei sehr großen Stichproben (hier: 20.481 Befragte im Datensatz) u.U. auch geeignete Modelle abgelehnt werden. Da sich im Fall des Datensatzes dieser Arbeit zeigt, dass auch der korrigierte χ^2 -Test nicht zur Messung der Modellgüte geeignet ist, wird nun die Güte der Modelle anhand der Fit-Indizes SRMR, CFI und TLI beurteilt. In Tabelle 4.2 sind die Werte dieser Fit-Indizes sowohl bei Verwendung der ML-Schätzmethode sowie in Klammern bei Verwendung der ADF-Schätz-

⁴⁰Mit der Skala ist hier das arithmetische Mittel aus den entsprechenden Itemwerten gemeint.

methode dargestellt. Bei den Konstrukten „Fachkompetenz“, „Diagnostik“, „Mindestanforderungen“, „Leistungserwartungen“, „Binnendifferenzierung“ und „Schulbeurteilung“ kann die Güte der Modelle auf Grund der zu geringen Anzahl an Informationen so nicht festgestellt werden, da es sich um gerade identifizierte Modelle handelt. Diese Messmodelle müssen daher im Weiteren noch gesondert untersucht werden. Auf Basis der Ergebnisse der überidentifizierten Messmodelle lässt sich feststellen, dass die Modellgüte bei Verwendung der ML-Schätzmethode grundsätzlich höher durch die Indizes beurteilt wird als bei der ADF-Schätzmethode. Außerdem haben Messmodelle mit mehr Indikatorvariablen eine tendenziell weniger hohe Güte. Basierend auf dem SRMR besitzen alle Modelle mindestens eine akzeptable Güte, da alle SRMR-Werte unter 0,08 liegen. Dieses Ergebnis muss aber erst noch durch den CFI und den TLI validiert werden.

Tab. 4.2: Beurteilung der Modellgüte bei Verwendung der ML-Methode
[ADF-Methode]

Konstrukt	SRMR	CFI	TLI
Fachkompetenz	–	–	–
Unterrichtsführung	0,055 [0,060]	0,903 [0,720]	0,854 [0,581]
Diagnostik	–	–	–
Sozialklima	0,025 [0,025]	0,981 [0,951]	0,943 [0,852]
Veranschaulichung	0,029 [0,028]	0,974 [0,940]	0,949 [0,881]
Interessantheit	0,006 [0,006]	0,999 [0,997]	0,997 [0,992]
Feedback	0,020 [0,020]	0,985 [0,964]	0,954 [0,892]
Materialien	0,016 [0,016]	0,993 [0,977]	0,978 [0,931]
Schüleraktivierung	0,008 [0,008]	0,998 [0,996]	0,994 [0,989]
Schülerautonomie	0,053 [0,057]	0,869 [0,725]	0,825 [0,633]
Mindestanforderungen	–	–	–
Leistungserwartungen	–	–	–
Binnendifferenzierung	–	–	–
Zeitmanagement	0,034 [0,034]	0,975 [0,935]	0,950 [0,871]
Selbstengagement	0,049 [0,049]	0,927 [0,890]	0,782 [0,669]
Leistungsfähigkeit	0,042 [0,043]	0,958 [0,925]	0,873 [0,776]
Interesse	0,020 [0,021]	0,992 [0,981]	0,977 [0,944]
Schulbeurteilung	–	–	–
Lernzuwachs	0,024 [0,026]	0,981 [0,949]	0,943 [0,848]

Der CFI, der mehr Parameter in die Berechnung einbezieht als der TLI, und der TLI beurteilen sowohl verschiedenen Modelle als auch gleiche Modell z.T. sehr unterschiedlich. Grundsätzlich lässt sich im Bezug auf den CFI und den TLI feststellen, dass der CFI unabhängig von der Schätzmethode höher als der TLI ist. In Tabelle 4.2 sind beim CFI und beim TLI Werte, die eine akzeptable Modellgüte ($> 0,95$) oder eine gute Modellgüte ($> 0,97$) anzeigen, hervorgehoben. Die Messmodelle für die Konstrukte „Unterrichtsführung“, „Schülerautonomie“ und „Selbstengagement“ besitzen demnach in allen Fällen weder eine gute noch eine akzeptable Modellgüte und die Messmodelle für die Konstrukte „Veranschaulichung“, „Leistungsfähigkeit“ und „Lernzuwachs“ besitzen überwiegend nicht einmal eine akzeptable Modellgüte, d.h. die beiden Fit-Indizes sind bei den beiden Schätzmethoden mehrheitlich unter 0,95. Bei den Konstrukten „Sozialklima“ und „Zeitmanagement“ kann keine eindeutig Aussage bezüglich der Modellgüte auf Basis des CFI und TLI getroffen werden, da keine eindeutige Richtung bezüglich der Modellgüte angezeigt wird. Bei den Konstrukten „Interessantheit“ und „Schüleraktivierung“ kann auf Basis dieser beiden Fit-Indizes von guten Modellen, bei den Konstrukten „Materialien“ und „Interesse“ von insgesamt guter Modellgüte und beim Konstrukt „Feedback“ von insgesamt noch akzeptable Modellgüte gesprochen werden.

Da sowohl bei den gerade identifizierten, als auch bei den überidentifizierten Messmodellen bei beiden Schätzmethoden die standardisierten partiellen Regressionsgewichte aller Indikatorvariablen positiv sind⁴¹ sowie unter eins liegen und die Standardfehler der geschätzten, nicht-standardisierten partiellen Regressionsgewichten, die die mögliche Streuung der jeweiligen Parameterschätzung angeben, alle deutlich kleiner als 0,05 sind, ist von verlässlichen Messungen auszugehen.⁴² Auf Grund dessen muss auf Basis der standardisierten partiellen Regressionsgewichte, welche bei einfaktoriellen Messmodellen

⁴¹Wären bei bestimmten Items keine Recodierungen durchgeführt worden, hätte man bei diesen Items negative standardisierten partiellen Regressionsgewichte erhalten. Damit zeigt sich auch, dass die Recodierung zur inhaltlichen Gleichrichtung der Antworten richtig war.

⁴²Gleiches gilt für die Standardfehler der Fehlervarianzen und bei beiden Metriken.

den Korrelationen zwischen den Indikatorvariablen und der latenten Variable entsprechen, bzw. der Faktorladungen und der standardisierten Fehlervarianzen der Items entschieden werden, ob sich alle in ein Modell einfließenden Items als Indikator für ein bestimmtes Konstrukt eignen. Die standardisierten Fehlervarianzen lassen sich durch die Differenz aus eins und den quadrierten Faktorladungen zwischen den latenten Variablen und den jeweiligen Items errechnen. Die quadrierten Faktorladungen bzw. sog. Kommunalitäten geben den Varianzanteil eines Items an, der durch die latente Variable erklärt wird, und demnach geben die standardisierten Fehlervarianzen den nicht erklärten Teil an. Ist die Korrelation zwischen einem Konstrukt und einer Indikatorvariablen größer als 0,707 so ist die Fehlervarianz kleiner als 0,5 und damit kann mehr als 50% der Varianz bei einem Item durch das jeweilige Konstrukt erklärt werden. Wenn bei Messmodellen mit nicht mindestens akzeptabler Güte die Faktorladungen und Fehlervarianzen gegen die Eignung bestimmter Items als Indikator für ein Konstrukt sprechen, weil die Faktorladungen relativ gering bzw. die Fehlervarianzen relativ groß sind, kann u.U. der Ausschluss solcher Items aus den Messmodellen zu einer Verbesserung der Modellgüte führen. Ein solcher Ausschluss muss allerdings auch inhaltlich begründbar sein (Backhaus et al. 2008A, S. 22-24 & 47-48, Bühner 2006, S. 288-290).

Zunächst werden bei den überidentifizierten Messmodellen mit nicht mindestens akzeptabler Güte die Faktorladungen und Fehlervarianzen der einzelnen Items untersucht, um mögliche Items ohne Eignung als Indikator für ein Konstrukt zu identifizieren und dann auf Basis dieser Erkenntnisse die Güte von modifizierte Messmodelle überprüfen zu können. Statt einer sehr restriktiven Faktorladung von 0,707 wird hier für die Eignung eines Items als Indikator eine Mindestfaktorladung von 0,6 festgesetzt.⁴³ Im Rahmen einer angestrebten Verbesserung der Modellgüte stehen demnach Items mit

⁴³Im Rahmen der Beurteilung einzelner Items als Indikator für ein Konstrukt wurde zunächst eine Mindestfaktorladung von 0,5 festgesetzt. Da sich bei der anschließenden Beurteilung von modifizierten Messmodellen ohne die identifizierten Items zeigte, dass eine Mindestfaktorladung von 0,5 nicht als adäquater „Cut-off“ geeignet ist, wird hier gleich von einer Mindestfaktorladung von 0,6 ausgegangen.

einer Faktorladung kleiner als 0,6 bei der Modifikation der Messmodelle zur Disposition, wobei auch sachlogischen Überlegungen in die Modifikationen einfließen. In Tabelle 4.3 sind alle Faktorladungen der entsprechenden Items sowie deren Fehlervarianzen dargestellt, wobei die Ergebnisse in Klammern wieder die ADF-Schätzmethode betreffen. Die Items eines Konstrukts wurden dabei nach der Größe der Faktorladungen geordnet.

Tab. 4.3: Ausgewählte Faktorladungen der überidentifizierten Messmodelle mit unzureichende Güte

Konstrukt	Item	Faktorladung		standardisierte Fehlervarianz	
Unterrichtsführung	ua05	0,477	[0,533]	0,772	[0,716]
	ua04	0,532	[0,572]	0,717	[0,673]
Sozialklima	uh01	0,567	[0,571]	0,679	[0,674]
Veranschaulichung	ub02	0,430	[0,448]	0,815	[0,799]
	ub01	0,572	[0,589]	0,673	[0,653]
	ub04	0,592	[0,601]	0,650	[0,639]
Schülerautonomie	lf05	0,298	[0,321]	0,911	[0,897]
	le03	0,383	[0,353]	0,583	[0,875]
	lf02	0,395	[0,443]	0,844	[0,804]
	lf03	0,414	[0,415]	0,829	[0,828]
	lf01	0,454	[0,466]	0,794	[0,783]
	le01	0,540	[0,518]	0,708	[0,732]
	le05	0,588	[0,655]	0,654	[0,571]
Zeitmanagement	ud04r	0,259	[0,261]	0,933	[0,932]
	ud05	0,376	[0,389]	0,859	[0,849]
	ud03	0,389	[0,403]	0,849	[0,838]
Selbstengagement	mc02r	0,372	[0,390]	0,862	[0,848]
	mc03	0,593	[0,638]	0,648	[0,593]
Leistungsfähigkeit	mb04r	0,502	[0,529]	0,748	[0,720]
	mb03	0,577	[0,613]	0,667	[0,624]
Lernzuwachs	za03	0,527	[0,543]	0,722	[0,705]

Grundsätzlich lässt sich bei fast allen dargestellten Items außer den Items „le01“ und „le03“ beim Konstrukt „Schülerautonomie“ feststellen, dass die Faktorladungen bei der ADF-Methode höher als bei der ML-Schätzmethode sind. Außerdem haben beim Konstrukt „Schülerautonomie“ sieben von insgesamt neun Indikatorvariablen eine Faktorladung unterhalb von 0,6 und somit

sind bei dem Messmodell dieses Konstrukts die mit Abstand meisten Items zu finden, die für eine unzureichende (nicht mindestens akzeptable) Modellgüte verantwortlich sein könnten und damit teilweise nicht als Indikator für das Konstrukt verstanden wurden.

Nachdem bei den überidentifizierten Modellen mit unzureichender Güte einige Items identifiziert werden konnten, die möglicherweise nicht als Indikator für ein Konstrukt geeignet sind, werden nun modifizierte Messmodelle, die die identifizierten Indikatorvariablen nicht mehr berücksichtigen, aufgestellt und im Bezug auf ihre Güte untersucht. Die Betrachtung der Modellgüte gelingt allerdings nur bei den modifizierten Messmodellen der Konstrukte „Unterrichtsführung“, „Veranschaulichung“, „Schülerautonomie“ und „Zeitmanagement“, da es sich bei diesen Modell auch nach Entfernung eines Items um überidentifizierte Messmodelle handelt. Deswegen werden die Modelle dieser Konstrukte als erstes betrachten und dann separat die nicht überidentifizierten Modelle. In Tabelle 4.4 sind die modifizierten Messmodelle sowie deren Güte bei beiden Schätzmethoden dargestellt, wobei die Messmodelle analog zu Tabelle 4.3 sortiert und Modellgüten größer oder gleich 0,95 wieder hervorgehoben wurden. Bei den Konstrukten „Veranschaulichung“ und „Zeitmanagement“ können durch die Entfernung der einzelnen identifizierten Items meist Verbesserungen der Modellgüte im Vergleich zum Ausgangsmodell erreicht werden, jedoch führt der Ausschluss des Items mit der niedrigsten Faktorladung bei einem Konstrukt nicht zwangsläufig zum Modell mit der höchsten Güte. Nachdem die verschiedenen Messmodelle geprüft wurden, muss nun unter Berücksichtigung der Sachlogik entschieden werden, welches Modell zur Abbildung eines Konstrukts am besten geeignet ist. Beim Konstrukt „Veranschaulichung“ kann der Ausschluss des Items „ub02“ sachlogisch begründet werden, denn es zielt im Gegensatz zu den anderen Items des Konstrukts als einziges Item wirklich auf die Veranschaulichung im Unterricht ab. Die anderen Items zielen hingegen auf die Verknüpfung von verschiedenen Lehrinhalte und deren Wichtigkeit im Bezug auf die Praxis ab. Auf Grund dessen wird das Konstrukt „*Veranschaulichung*“ in seiner modifizierten Form (ohne das Item „ub02“) in das Konstrukt „Thematische

Tab. 4.4: Modifikationen der Messmodelle und deren Güte bei Verwendung der ML-Methode [ADF-Methode]

Konstrukt	SRMR	CFI	TLI
• Unterrichtsführung			
- Ausgangsmodell	0,055 [0,060]	0,903 [0,720]	0,854 [0,581]
- ohne ua01	0,045 [0,050]	0,930 [0,841]	0,884 [0,690]
- ohne ua02	0,036 [0,039]	0,951 [0,841]	0,918 [0,735]
- ohne ua03	0,055 [0,059]	0,893 [0,751]	0,822 [0,584]
- ohne ua01 & ua02	0,030 [0,032]	0,969 [0,914]	0,937 [0,828]
- ohne ua01 & ua03	0,025 [0,026]	0,979 [0,942]	0,958 [0,884]
- ohne ua02 & ua03	0,021 [0,021]	0,983 [0,956]	0,966 [0,912]
- ohne ua01 - ua03	0,022 [0,023]	0,982 [0,960]	0,945 [0,881]
• Veranschaulichung			
- Ausgangsmodell	0,029 [0,028]	0,974 [0,940]	0,949 [0,881]
- ohne ub02	0,020 [0,019]	0,989 [0,977]	0,968 [0,932]
- ohne ub01	0,026 [0,025]	0,984 [0,962]	0,951 [0,886]
- ohne ub04	0,029 [0,028]	0,979 [0,955]	0,936 [0,865]
• Schülerautonomie			
- Ausgangsmodell	0,053 [0,057]	0,869 [0,725]	0,825 [0,633]
- le01 - le05	0,056 [0,058]	0,907 [0,836]	0,814 [0,671]
- lf01 - lf03, lf05	0,025 [0,025]	0,955 [0,936]	0,866 [0,809]
• Zeitmanagement			
- Ausgangsmodell	0,034 [0,034]	0,975 [0,935]	0,950 [0,871]
- ohne ud04r	0,040 [0,039]	0,976 [0,939]	0,928 [0,818]
- ohne ud05	0,008 [0,008]	0,998 [0,993]	0,993 [0,980]
- ohne ud03	0,009 [0,009]	0,999 [0,996]	0,996 [0,988]

Verknüpfung“ umbenannt und man kann nun auf Basis der Ergebnisse der KFA und der Sachlogik davon ausgehen, dass dieses Konstrukt adäquat abgebildet werden konnte. Beim Konstrukte „Klassen- und *Zeitmanagement*“ soll auf Basis der Modellgüte der modifizierten Modelle entweder das Item „ud05“ oder das Item „ud03“ aus dem ursprünglichen Messmodell entfernt werden. Das Item „ud03“ zielt auf die Beteiligung der Schülerinnen und Schüler am Unterricht und das Item „ud05“ auf das Zeitmanagement der Lehrerin oder des Lehrers ab, wohingegen die restlichen Items auf das Klassenmanagement des Lehrers anspielen. Aus sachlogischen Überlegungen bietet sich daher beim Konstrukt „Klassen- und *Zeitmanagement*“ an, dass Item „ud03“ zu entfer-

nen. Das Konstrukt kann damit auch weiterhin „Klassen- und *Zeitmanagement*“ genannt werden und wurde demnach auch so von den Schülerinnen und Schülern verstanden.

Bei dem Konstrukt „Unterrichtsführung“ stehen auf Basis der Faktorladungen die Items „ua05“ und „ua04“ zur Disposition. Werden diese Items einzeln oder zusammen aus dem Messmodell entfernt verbessert sich zwar die Modellgüte, aber es wird trotzdem keine mindestens akzeptable Modellgüte erreicht. Auf Grund dessen ist die bisherige Verfahrensweise beim Konstrukt „Unterrichtsführung“ nicht zielführend. Betrachtet man deshalb nun alle Items des Konstrukts „*Unterrichtsführung* und *Strukturiertheit*“ lassen sich verschiedene thematische Ausrichtungen erkennen. Die Items „ua04“ bis „ua07“ sind klar auf die *Strukturiertheit* des Unterrichts ausgerichtet (siehe Anhang B). Hingegen zielen die Items „ua01“ bis „ua03“ auf die Verständlichkeit und Klarheit des Ausdrucks und der Anweisungen der Lehrerin oder des Lehrers ab. Auf Grund dessen wurden im Rahmen der Modellmodifikationen beim Konstrukt „Unterrichtsführung“ in einem zweiten Schritt die Items „ua01“, „ua02“ und „ua03“ einzeln und in verschiedenen Kombinationen zur Disposition gestellt. In Tabelle 4.4 ist ersichtlich, dass das Messmodell des Konstrukts nur eine insgesamt akzeptable Modellgüte besitzt, wenn die Items „ua01“ sowie „ua04“ bis „ua07“ berücksichtigt werden. Somit wurden neben den Items „ua04“ bis „ua07“ auch das Item „ua01“ als Indikator für die Führung des Unterrichts durch den Lehrer und die *Strukturiertheit* seines Unterrichts verstanden, wohingegen die beiden verbleibenden Items nicht als dem Konstrukt zugehörig angesehen wurden.

Bei der Konzeption des Fragebogens waren die Items des Konstrukts „*Schülermitbestimmung* und *-autonomie*“ ursprünglich zur Abbildung von zwei verschiedenen Konstrukten gedacht. So sollte einerseits das Konstrukt „Schülermitbestimmung“ durch die Items „le01“ bis „le05“ und andererseits das Konstrukt „Schülerautonomie“ durch die Items „lf01“ bis „lf05“ abgebildet werden. Allerdings konnte im Rahmen der Untersuchungen von Wagner (2008) keine zufriedenstellende innere Konsistenz gemessen durch den Cronbach- α -Konsistenzkoeffizienten bei den getrennten Konstrukten festge-

stellt werden, weshalb diese beiden ursprünglichen Einzelkonstrukte zusammengefasst wurden. Auf Basis dieser Erkenntnis wurde auch im Rahmen dieser Arbeit ein gemeinsames Konstrukt „*Schülermitbestimmung* und -*autonomie*“ unterstellt. Genau wie für die Teilkonstrukte „Schülermitbestimmung“ und „Schülerautonomie“ kann auch für das Gesamtkonstrukt „*Schülermitbestimmung* und -*autonomie*“ keine insgesamt mindestens akzeptable Güte des Messmodells festgestellt werden (siehe Tabelle 4.4). Auf Basis der Faktorladungen kommen beim Konstrukt „*Schülermitbestimmung* und -*autonomie*“ wie bereits erwähnt sieben von neun Items zur Entfernung aus dem Messmodell in Betracht. Deshalb wurde nach und nach das Item mit der jeweils niedrigsten Faktorladung aus dem Messmodell des Gesamtkonstrukts entfernt. Durch die kumulative Entfernung der Items „lf05“, „lf02“, „lf03“ und „le03“ konnten jeweils Verbesserungen der Modellgüte, gemessen durch den CFI und den TLI, erreicht werden (siehe Tabelle 4.5). Allerdings konnte dann durch die anschließende Entfernung des Items mit der niedrigsten Faktorladung („lf01“) keine Verbesserung mehr erreicht werden. Daher wurde im nächsten Schritt statt des Items „lf01“ das Item „le01“, das die zweitniedrigste Faktorladung im entsprechenden Modell⁴⁴ besitzt, entfernt und damit eine insgesamt akzeptable Modellgüte erreicht werden.

Tab. 4.5: Modifikationen des Messmodells „*Schülermitbestimmung* und -*autonomie*“ und deren Güte bei Verwendung der ML-Methode [ADF-Methode]

Modell	SRMR	CFI	TLI
(0) Ausgangsmodell	0,053 [0,057]	0,869 [0,725]	0,825 [0,633]
(1) ohne lf05	0,054 [0,058]	0,876 [0,745]	0,827 [0,643]
(2) ... & ohne lf02	0,045 [0,049]	0,910 [0,800]	0,866 [0,700]
(3) ... & ohne lf03	0,046 [0,049]	0,918 [0,836]	0,863 [0,726]
(4) ... & ohne le03	0,029 [0,030]	0,964 [0,920]	0,929 [0,840]
(5a) ... & ohne lf01	0,034 [0,034]	0,963 [0,927]	0,888 [0,782]
(5b) ... & ohne le01	0,021 [0,021]	0,984 [0,966]	0,951 [0,898]

⁴⁴Mit dem entsprechenden Messmodell ist das aus den folgenden Items bestehende Modell gemeint: „le01“, „le02“, „le04“, „le05“ sowie „lf01“.

Einerseits konnte ein Modell⁴⁵ mit zufriedenstellender Güte gefunden werden, andererseits muss das Modell nun auch einer sachlogischen Diskussion standhalten. In Tabelle 4.6 sind die Items des hier unterstellten Gesamtkonstrukts „Schülerautonomie“ sowohl deren Thematiken dargestellt. Die Items „le01“ und „le03“ besitzen wie im Rahmen der KFA festgestellt keine ausreichende Reliabilität mit dem Konstrukt. Dies kann sachlogisch damit erklärt werden, dass die beiden Items wahrscheinlich nicht wie von den Fragebogenkonstrukteuren gedacht als Mitbestimmung durch die Schülerinnen und Schüler, sondern vielmehr als Einbindung der Schülerinnen und Schüler in den Unterricht von den Befragten verstanden wurden bzw. auf die gemeinsame Diskussionskultur ausgerichtet sind. Im Gegensatz zu den Items „le01“

Tab. 4.6: Items und Themenbereiche des Konstrukts
„Schülermitbestimmung und -autonomie“

Item	Thematik
le01 Meine Lehrerin/mein Lehrer und die Klasse diskutieren gemeinsam.	<u>Schülereinbindung</u> im Unterricht
le03 Meine Lehrerin/mein Lehrer lässt uns im Unterricht Fragen stellen.	
le02 Meine Lehrerin/mein Lehrer ermöglicht es uns, eigene Vorschläge in die Unterrichtsgestaltung einzubringen.	<u>Schülermitbestimmung</u> bei der Unterrichtsgestaltung (aktiv/passiv)
le04 Meine Lehrerin/mein Lehrer nimmt Beiträge und Ideen von uns Schülern/Schülerinnen in den Unterricht auf.	
le05 Meine Lehrerin/mein Lehrer lässt uns einzelne Unterrichtsstunden mitgestalten.	
lf01 Der Lehrer/die Lehrerin lässt uns eigene Lösungswege finden.	<u>Schülerautonomie</u> bei der Aufgabenbewältigung ... bei Diskussionen ... bei der Arbeitseinteilung ... bei der Beurteilung
lf02 Die Gesprächsleitung haben manchmal die Schüler.	
lf03 Bei umfangreichen Aufgaben kann ich mir die Arbeit selbst einteilen.	
lf05 Wir beurteilen uns gegenseitig.	

⁴⁵Die Prüfung von anderen Kombinationen der Items sowie Messmodellen aus den verbleibenden Items führten zu keiner insgesamt mindestens akzeptablen Modellgüte oder zu Faktorladungen von größer oder gleich 0,6 bei allen Items.

und „le03“ kann bei den Items „le02“, „le04“ und „le05“, die klar nach dem Bestehen von Gestaltungsmöglichkeiten im Unterricht durch die Schülerinnen und Schüler fragen, ein mitbestimmender Charakter im klassischen Sinn erkannt werden.

Weiterhin zielen die Items mit der Kennung „lf“ zwar alle auf die Autonomie der Schülerinnen und Schüler im Unterricht ab, allerdings bestehen bei diesen Items qualitative Unterschiede. Beispielsweise ist bei den Items „lf02“, „lf03“ und „lf05“ fraglich, ob die jeweiligen Aspekte der Autonomie als von Schülerinnen und Schülern bestimmte Vorgehensweisen im Unterricht von den Befragten erkannt bzw. wahrgenommen werden. Die Gründe dafür könnten sein, dass diese Form der Autonomie zu selten gewährt und nur nach der vorherigen Anweisung durch den Lehrenden durchgeführt wird. Damit verlieren die angesprochenen Unterrichtsmethoden teilweise ihren durch Schülerentscheidungen beeinflussten Charakter. Speziell beim Item „lf05“ ist zweifelhaft, ob eine gegenseitige Beurteilung von jeder Schülerin bzw. jedem Schüler als autonome Handlung bzw. Handlung mit Auswirkungen wahrgenommen wird, wenn die gegenseitige Beurteilung nur als Anregung von den Klassenkameraden verstanden wird bzw. die Schülerinnen und Schüler nicht das Gefühl haben, dass die gegenseitige Beurteilung auch die Beurteilung durch den Lehrer beeinflusst. Hingegen scheinen die Befragten beim Item „lf01“ die eigene Autonomie im Bezug auf die Aufgabenbewältigung insgesamt verstanden zu haben bzw. zu kennen und den Zusammenhang mit den Items „le02“, „le04“ und „le05“ erkannt zu haben. Insgesamt sind die Items dieses Konstrukts auf Grund der verschiedenen Teilaspekte und den damit verbundenen Problemen als relativ heterogen einzuschätzen und konnten deshalb nicht in ihrer Gesamtheit als Messmodell für das Konstrukt „*Schülermitbestimmung und -autonomie*“ verwendet werden. Trotzdem konnte letztendlich ein statistisch wie sachlogisch akzeptables Messmodell gefunden werden, dass die sich bedingenden Aspekte der Mitbestimmung und Autonomie von Schülerinnen und Schülern im Unterricht verlässlich abbildet.

Bei den ursprünglich überidentifizierten Messmodellen der Konstrukte „Sozialklima“, „Selbstengagement“, „Leistungsfähigkeit“ und „Lernzuwachs“ führt

wie bereits erwähnt die Modifikation der Messmodelle bzw. Entfernung einer Indikatorvariablen zu gerade identifizierten Messmodellen, bei denen eine Prüfung der Modellgüte nicht möglich ist. Auf Grund dessen muss auf Basis der Faktorladungen bzw. Fehlervarianzen der Items bei den modifizierten Modelle und der Sachlogik entschieden werden, ob durch die Entfernung der identifizierten Items (siehe Tabelle 4.3) eine verbesserte Abbildung der Konstrukte zu erwarten ist. In Tabelle 4.7 sind neben den nach der Größe geordneten Faktorladungen auch die Fehlervarianzen dieser Konstrukte dargestellt. Da es bei diesen Modellen um gerade identifizierten Messmodellen handelt und diese nur eine Lösung für die Parameter besitzen, kommen die ML- und ADF-Schätzmethode zu den gleichen Ergebnissen und werden deswegen nicht gesondert ausgewiesen.

Tab. 4.7: Faktorladungen der gerade identifizierten, modifizierten Messmodelle

Konstrukt	Item	Faktorladung	standardisierte Fehlervarianz
Sozialklima	uh02	0,663	0,560
	uh04	0,722	0,479
	uh03	0,747	0,442
Selbstengagement	mc03	0,539	0,709
	mc04	0,651	0,576
	mc01	0,752	0,434
	mc02r	0,293	0,914
	mc04	0,681	0,536
	mc01	0,719	0,483
Leistungsfähigkeit	mb04r	0,618	0,618
	mb03	0,652	0,575
	mb01	0,665	0,558
Lernzuwachs	za02	0,662	0,562
	za04	0,707	0,500
	za01	0,823	0,323

Beim Konstrukt „Sozialklima“ konnte das Item „uh01“ als möglicherweise nicht geeigneter Indikator identifiziert werden. Betrachtet man deshalb die Faktorladungen des modifizierten Messmodells sieht man in Tabelle 4.7, dass

bei Nichtberücksichtigung des Items „uh01“ die Faktorladungen aller verbleibender Items über 0,6 sind. Das Item „uh01“ fragt danach, ob eine Lehrerin oder ein Lehrer Probleme in der Klasse bemerkt. Dagegen wird mit den Items „uh02“ bis „uh04“ danach gefragt, ob der Lehrer eine Vertrauensperson ist, er sich Zeit für die Schülerinnen und Schüler nimmt bzw. ob er sich der Probleme der Schüler annimmt. Ausgehend von den Fragestellungen und den Faktorladungen ist das Item „uh01“ nicht zur Abbildung des Konstrukts „Sozialklima“ geeignet, da das Bemerken von Problemen im Gegensatz zu den Items „uh02“ bis „uh04“ nicht als Aspekt des Lehrer-Schüler-Sozialklimas von den Schülerinnen und Schülern verstanden wurde. Auch beim Konstrukt „Lernzuwachs“ konnte ein Items („za03“) mit möglicherweise mangelnder Eignung als Indikator identifiziert werden. Bei dem modifizierten Messmodell mit den Items „za01“, „za02“ und „za04“ sind ausschließlich Faktorladungen über 0,6 geschätzt worden. Damit kann davon ausgegangen werden, dass das Item „za03“, dass nach dem Erlernen von Verknüpfungen zwischen Sachgebieten fragt, von den Schülerinnen und Schülern nicht im Kontext mit den Items „za01“, „za02“ und „za04“, die nach dem Lernzuwachs im letzten Halbjahr und im Bezug auf den durchführenden Lehrer sowie nach der Verwendbarkeit des Gelernten fragen, gesehen wurde. Da bei den modifizierten Messmodellen der Konstrukte „Sozialklima“ und „Lernzuwachs“ alle Faktorladungen größer als 0,6 sind, werden diese Modelle als zur Abbildung der Konstrukte geeignet angesehen.

Beim ursprünglichen Messmodell des Konstrukts „Selbstengagement“ wurden die Items „mc02r“ und „mc03“ identifiziert und deshalb einzeln bei der Modifikation nicht mehr berücksichtigt. Das modifizierte Messmodell aus den Items „mc01“, „mc03“ und „mc04“ ist auf Basis der Faktorladungen dem Messmodell aus den Items „mc01“, „mc02r“ und „mc04“ vorzuziehen (siehe Tabelle 4.7).⁴⁶ Damit ist das Item „mc02r“ am wenigsten zur Abbildung des Konstrukts „Selbstengagement“ geeignet. Sachlogisch kann die mangelnde Eignung des Items „mc02r“ damit begründet werden, dass es im Gegensatz

⁴⁶Modifizierte Messmodelle aus den Items „mc02r“, „mc03“ und „mc04“ bzw. „mc01“, „mc02r“ und „mc03“ eignen sich auf Basis der Faktorladungen im Vergleich zu den beiden im Text beschriebenen Messmodellen in keinem Fall.

zu den anderen Items nicht auf das eigene Engagement des Schülers für die Schule abzielt, sondern ob eine Schülerin oder ein Schüler sich leicht ablenken lässt. Wenn sich ein Schüler leicht ablenken lässt, muss dies aus Schülersicht allerdings noch nicht bedeuten, dass eine Schülerin oder ein Schüler nicht auch im Rahmen seiner Möglichkeiten ein unterrichtliches Selbstengagement betreibt. Zwar hat beim zu präferierenden Messmodell das Item „mb03“ eine Faktorladung kleiner als 0,6 (0,551), jedoch wird im Weiteren dieses Messmodell zugrundegelegt, da es die insgesamt höchsten einzelnen Faktorladungen besitzt und auf Grund der Eigenschaft von null Freiheitsgraden nicht im Bezug auf die Modellgüte untersucht werden kann. Außerdem würde eine weitere Reduktion des Messmodells zu einem unteridentifizierten Modell führen, dass im Rahmen der KFA nicht ohne Restriktionen geschätzt werden könnte. Im Rahmen der in Teil 4.2 folgenden Analysen wird sich dann zeigen müssen, ob die Abbildung des Konstrukts „Selbstengagement“ wirklich gelungen ist bzw. ob auf Basis dieses Messmodells weitergehende Analysen zu nachvollziehbaren Ergebnissen führen.

Auf Basis der bisherigen Erkenntnisse sollten beim Konstrukt „Leistungsfähigkeit“ die Items „mb04r“ und „mb03“ im Rahmen einer Modellverbesserung zur Disposition gestellt werden. Allerdings wurden in diesen modifizierten Messmodellen auch Faktorladungen kleiner als 0,6 geschätzt. Auf Grund dessen wurden alle Kombinationen aus drei der vier Items des Konstrukts im Hinblick auf die Faktorladungen untersucht. In Tabelle 4.7 sind nur die Faktorladungen bei dem Messmodell mit den Items „mb01“, „mb03“ und „mb04r“ dargestellt, da dabei die Items als einziges alle eine Faktorladung größer als 0,6 haben. Betrachtet man die Formulierungen der Items beim Konstrukt „Leistungsfähigkeit“ fällt auf, dass das Item „mb02“ speziell auf die Leistungsfähigkeit der Schülerinnen und Schüler in Klassenarbeiten abzielt. Die restlichen Items fragen hingegen nach der generellen Leistungsfähigkeit sowie der Fähigkeit Lerninhalte zu verstehen und zu lernen. Auch besitzt das Item „mb02“ eine andere Qualität, da im Gegensatz zu den anderen Items die Beantwortung dieses Items weniger von den Wahrnehmungen der Schülerinnen und Schüler abhängen sollte und mehr von dem Faktum

der geschriebenen Noten in Klassenarbeiten, und es wird deshalb im Weiteren nicht mehr beim Konstrukt „Leistungsfähigkeit“ berücksichtigt.

Abschließend werden nun die Faktorladungen und standardisierten Fehlervarianzen der ursprünglich gerade identifizierten Messmodelle der Konstrukte „Fachkompetenz“, „Diagnostik“, „Mindestanforderungen“, „Leistungserwartungen“, „Binnendifferenzierung“ und „Schulbeurteilung“ untersucht und auf Basis derer versucht die Güte dieser Modelle zu beurteilen.

Tab. 4.8: Faktorladungen der gerade identifizierten, ursprünglichen Messmodelle

Konstrukt	Item	Faktorladung	standardisierte Fehlervarianz
Fachkompetenz	ui03	0,619	0,617
	ui01	0,653	0,574
	ui02	0,787	0,381
Diagnostik	ug02	0,641	0,589
	ug03	0,658	0,567
	ug01	0,770	0,407
Mindestanforderungen	uf02	0,664	0,559
	uf03	0,758	0,425
	uf01	0,819	0,329
Leistungserwartungen	ue03	0,365	0,867
	ue02	0,671	0,550
	ue01	0,824	0,321
Binnendifferenzierung	lb03	0,564	0,682
	lb02	0,677	0,542
	lb01	0,797	0,365
Schulbeurteilung	ga01	0,807	0,349
	ga02	0,865	0,252
	ga03	0,894	0,201

Auf Basis der Faktorladungen und standardisierten Fehlervarianzen kann bei den Konstrukten „Fachkompetenz“, „Diagnostik“, „Mindestanforderungen“ und „Schulbeurteilung“ von adäquaten Messungen ausgegangen werden, da bei diesen Konstrukten alle Faktorladungen größer als 0,6 sind (siehe Tabelle 4.8). Beim Konstrukt „Binnendifferenzierung“ besitzt nur das Item „lb03“ eine Faktorladung kleiner als 0,6. Da allerdings auf Grund eines gerade iden-

tifizierten Messmodells keine abschließende Beurteilung der Modellgüte beispielsweise auf Basis der Fit-Indizes möglich ist und die kleinste Faktorladung zwischen 0,5 und 0,6 liegt, wird auch im Weiteren davon ausgegangen, dass das Konstrukt „Binnendifferenzierung“ durch die Indikatorvariablen „lb01“ bis „lb03“ bestimmt wird, und damit analog zur Argumentation beim Konstrukt „Selbstengagement“ verfahren. Außerdem ist es auch auf Basis der Sachlogik schwierig ein Item aus dem Messmodell zu entfernen. Zwar zielen die Items „lb01“ und „lb02“ auf die Differenzierung der Schwierigkeit von Aufgaben bei Schülerinnen und Schülern mit unterschiedlichen Voraussetzungen ab und das Item „lb03“ auf die Ausgewogenheit der Zusammensetzung von Arbeitsgruppen, jedoch würden durch die Entfernung des Items mit der kleinsten Faktorladung („lb03“) unterschiedliche Aspekte der Binnendifferenzierung keine Berücksichtigung mehr finden.

Beim Konstrukt „Leistungserwartungen“ besitzt das Item „ue03“ eine Faktorladung deutlich unter 0,6. Zwar kann im Rahmen der KFA ein Messmodell aus zwei Indikatoren ohne Restriktionen nicht geschätzt werden, allerdings besitzt das Item „ue03“ eine andere Ausrichtung als die Items „ue01“ und „ue02“. Denn während das Item „ue03“ auf die Schwierigkeit von Klassenarbeiten abzielt, sind die Items „ue01“ sowie „ue02“ auf die allgemeinen Leistungserwartungen ausgerichtet. Wie beim Konstrukt „Leistungsfähigkeit“ zeigt sich auch beim Konstrukt „Leistungserwartungen“, dass Items im Bezug auf Klassenarbeiten von Schülerinnen und Schülern anders verstanden bzw. wahrgenommen werden als allgemeiner gehaltene Items. Im Folgenden wird daher das Konstrukt „Unangemessenheit der *Leistungserwartungen*“ nur noch durch die Items „ue01“ und „ue02“ gemessen. Das Messmodell dieses Konstrukts ist damit unteridentifiziert und grundsätzlich im Rahmen der KFA nicht ohne weiteres lösbar, aber diese Problematik kann bei SGM durch die Einbeziehung weiterer Messmodelle bzw. mehrerer Konstrukte behoben werden. Natürlich muss sich bei den folgenden Analysen erst zeigen, ob auf Basis dieses Messmodells weitere Erkenntnisse gewonnen werden können.

Nachdem die Güte der überidentifizierten Messmodelle und die Faktorladungen der nicht-überidentifizierten Messmodelle unter Berücksichtigung der

Sachlogik überprüft wurden, kann man insgesamt von einer gelungenen Beobachtung der gewählten Konstrukte ausgehen. Zwar mussten leichte Modifikationen der Messmodelle durchgeführt werden, trotzdem ist bei den meisten der letztendlich zur Anwendung kommenden Messmodellen entweder die Modellgüte insgesamt mindestens akzeptabel oder die Faktorladungen der einzelnen Items sind alle größer als 0,6. In Anhang C sind nochmals die Konstrukte und deren Items, die auf Basis der bisherigen Erkenntnisse zusammengestellt wurden und nun im Weiteren bei den Strukturgleichungsmodellen analysiert werden, dargestellt. Die aussortierten Items werden weder zu neuen Messmodellen zusammengefügt noch zu bestehenden Messmodellen hinzugefügt, da dies sachlogisch nicht gerechtfertigt wäre oder die Messmodelle schon als mindestens akzeptabel angesehen werden können.

4.2 Analyse der Konstrukte

Nachdem im ersten Teil der Datenanalyse mit Hilfe der KFA geprüft wurde, ob anhand des verwendeten Fragebogens die Beobachtung der gewählten Konstrukte gelungen ist bzw. durch welche Items die Konstrukte am besten abgebildet werden können, wird nun im zweiten Teil der Datenanalyse mit Hilfe von Strukturgleichungsmodellen geprüft, in wie weit auf Basis von theoretischen und sachlogisch fundierten Überlegen angenommene Abhängigkeiten zwischen den hypothetischen Konstrukten empirisch bestätigt werden können. Dabei liegt der Fokus vor allem auf der Analyse der Einflussfaktoren auf den Erfolg von Schule bzw. die Unterrichtswirksamkeit gemessen durch den Lernzuwachs. Neben den Konstrukten, die durch die in Abschnitt 4.1 formulierten Messmodelle bestimmt werden (siehe Anhang C – Modifizierte Item-Konstrukt-Zuordnung auf Basis der KFA-Ergebnisse), fließen dabei auch Merkmale wie das Geschlecht, der Bildungsgang, die Klassenstärke und der Notenschnitt in Kategorien in die Analyse ein und es werden daher auch implizit Unterschiede zwischen verschiedenen Schülergruppen untersucht.

Als erstes wird ein Strukturgleichungsmodell (Modell 1) angenommen bei

dem das Konstrukt „Lernzuwachs“ als endogene Variable und die verbleibenden achtzehn Konstrukte⁴⁷ als exogene Variablen einfließen und damit unterstellt, dass der wahrgenommene Lernzuwachs der Schülerinnen und Schüler durch die einbezogenen exogenen latenten Variablen der empfundenen Qualität von Schulunterricht erklärt werden kann. Weiterhin wird angenommen, dass die Konstrukte „Unangemessenheit der *Leistungsanforderungen*“ und „Unangemessenheit der *Leistungserwartungen*“ den wahrgenommenen Lernzuwachs negativ beeinflussen, d.h. je unangemessener die Leistungsanforderungen bzw. -erwartungen empfunden werden, desto schlechter wird auch der Lernzuwachs wahrgenommen, und dass die anderen exogenen Variablen einen positiven Einfluss auf den Lernzuwachs haben. Außerdem wird von Korrelationen zwischen allen exogenen latenten Variablen und einem Messfehler im Bezug auf die endogene latente Variable „Lernzuwachs“ ausgegangen. Alle in das Modell einfließenden latenten Variablen werden durch die in Abschnitt 4.1 letztendlich unterstellten Messmodelle (siehe Anhang C für die unterstellte Item-Konstrukt-Zuordnung) operationalisiert, wobei wieder die Indikatorvariable als Referenzvariable in einem Messmodell angesehen wird, die die höchste korrigierte Korrelation mit der Skala aus den restlichen Indikatorvariablen des entsprechenden Konstrukts hat. Mit dieser Formulierung des SGM ist die Identifizierbarkeit der Modellstruktur sowie -güte gewährleistet und entsprechend den Ausführungen bei der KFA werden nun die Parameter mit Hilfe der ML- sowie ADF-Methode geschätzt.

Sowohl bei Verwendung der ML-Methode als auch bei Verwendung der ADF-Methode sind alle Standardfehler der geschätzten Parameter kleiner oder gleich 0,05 und es ist deshalb auf Basis dieser Werte von zuverlässigen Schätzungen der Parameter auszugehen. Weiterhin wird auf Grund der Ergebnisse der KFA aus Abschnitt 4.1 davon ausgegangen, dass die verwendeten Indikatorvariablen zur Abbildung der latenten Variablen geeignet sind. Die im

⁴⁷Exogene Variablen im Modell: „Fachkompetenz“, „Unterrichtsführung“, „Diagnostik“, „Sozialklima“, „Verknüpfung“, „Interessantheit“, „Feedback“, „Materialien“, „Schüleraktivierung“, „Schülerautonomie“, „Mindestanforderungen“, „Leistungserwartungen“, „Binnendifferenzierung“, „Zeitmanagement“, „Selbstengagement“, „Leistungsfähigkeit“, „Interesse“ und „Schulbeurteilung“.

Rahmen des SGM geschätzten Indikatorreliabilitäten liegen außerdem zwischen null und eins und es kann auch deswegen von der Zuverlässigkeit der Schätzungen ausgegangen werden. Betrachtet man in Tabelle 4.9 die Ergebnisse der Schätzung bei Verwendung der ML-Methode kann auf Basis der C.R., die sich aus der Division eines geschätzten unstandardisierten Parameterwerts und dessen Standardfehler der Schätzung ergeben, für die markierten Konstrukte „Sozialklima“, „Feedback“, „Schüleraktivierung“, „Schülerautonomie“, „Binnendifferenzierung“ und „Selbstengagement“ auf einem 95%-Niveau nicht ausgeschlossen werden, dass die geschätzten Parameter signifikant von null verschieden sind. Weiterhin besitzen die standardisierten Regressionsgewichte der markierten Konstrukte „Unterrichtsführung“, „Schüleraktivierung“ und „Schülerautonomie“ nicht wie angenommen einen positiven Einfluss und das Konstrukte „Leistungserwartungen“ nicht wie angenommen einen negativen Einfluss auf den wahrgenommenen Lernzuwachs. Einerseits wird durch die exogenen latenten Variablen insgesamt 80% der Variation des „Lernzuwachs“ erklärt und die Modellgüte ist auf Basis des SRMR als akzeptable anzusehen, andererseits zeigen der CFI und der TLI keine mindestens akzeptable Modellgüte an und die residualen Kovarianzen, die sich aus der Differenz zwischen der empirischen und modelltheoretischen Kovarianzmatrix berechnen, sind in mehreren Fällen deutlich über 0,1. Auf Grund der erhöhten residualen Kovarianzen ist davon auszugehen, dass eine unzureichende Modellanpassung vorliegt. Neben den bisher beschriebenen Auffälligkeiten bei den Parametern sind mehrere Korrelationen zwischen den exogenen latenten Variablen über 0,5 geschätzt worden und es ist auch deshalb davon auszugehen, dass Multikollinearität vorliegt. Bei der Betrachtung der Ergebnisse der ADF-Methode in Tabelle 4.9 (die Ergebnisse der Parameterschätzung mit der ADF-Methode sind wieder in Klammern dargestellt) fällt auf, dass die Probleme im Bezug auf die Schätzungen zwar ähnlich sind, aber auf Basis der C.R. werden nicht die komplett gleichen exogenen latenten Variablen als signifikant von null verschieden angesehen. Auch wird nicht den komplett gleichen Hypothesen bezüglich der Richtung der Zusammenhänge zwischen der endogenen und den exogenen latenten Variablen widersprochen,

Tab. 4.9: Ergebnisse der Parameterschätzung des SGM bei Verwendung der ML-Methode [ADF-Methode] – Modell 1

exogenes Konstrukt	C.R. der geschätzten Parameter		standardisiertes Regressionsgewicht	
Fachkompetenz	6,638	[0,442]	0,152	[0,012]
Unterrichtsführung	-8,329	[0,460]	-0,303	[0,021]
Diagnostik	5,963	[1,880]	0,107	[0,037]
Sozialklima	1,367	[5,221]	0,034	[0,133]
Verknüpfung	10,262	[3,671]	0,218	[0,125]
Interessantheit	2,000	[3,247]	0,051	[0,132]
Feedback	1,514	[3,497]	0,027	[0,081]
Materialien	21,338	[5,941]	0,333	[0,150]
Schüleraktivierung	-1,936	[-1,460]	-0,025	[-0,021]
Schülerautonomie	-0,395	[-3,416]	-0,010	[-0,085]
Mindestanforderungen	-9,100	[-1,260]	-0,088	[-0,018]
Leistungserwartungen	5,902	[6,357]	0,072	[0,109]
Binnendifferenzierung	0,529	[3,877]	0,005	[0,032]
Zeitmanagement	4,160	[-1,909]	0,049	[-0,036]
Selbstengagement	1,933	[-2,132]	0,019	[-0,031]
Leistungsfähigkeit	12,036	[11,651]	0,154	[0,151]
Interesse	28,478	[24,371]	0,352	[0,476]
Schulbeurteilung	6,733	[2,680]	0,046	[0,020]
Reliabilität der endogenen latenten Variable: 0,800 [0,854]				
SRMR: 0,046 [0,061] CFI: 0,895 [0,620] TLI: 0,881 [0,572]				

denn im Gegensatz zu den Ergebnissen bei Verwendung der ML-Methode wird der positive Einfluss der „Unterrichtsführung“ bestätigt und die positiven Einflüsse der Konstrukte „Zeitmanagement“ und „Selbstengagement“ nicht bestätigt. Die Modellgüte wird auf Basis des CFI und TLI bei Verwendung der ADF-Methode zudem deutlich schlechter beurteilt als bei der ML-Methode. Letztendlich können bei beiden Schätzmethoden die theoretisch aufgestellten Beziehungen empirisch nicht bestätigt werden. Da jedoch die Vermutung naheliegt, dass das aufgestellte Modell auf Grund von Multikollinearität zwischen den exogenen latenten Variablen nicht bestätigt werden kann, und weiterhin das Bestreben besteht die Einflussfaktoren auf den

wahrgenommenen Lernzuwachs adäquat zu identifizieren, wird das Strukturgleichungsmodell 1 modifiziert. Bei beiden Schätzmethoden besitzt das Konstrukt „Interesse“ das mit Abstand höchste standardisierte Regressionsgewicht und trägt damit am meisten zur Erklärung der Variation des Lernzuwachses bei. Daher werden nun zur Identifikation bzw. Beseitigung von Multikollinearität die Korrelationen zwischen der Variablen „Interesse“ und den weiteren exogenen latenten Variablen untersucht und auffällige Werte im Rahmen der Modellmodifikationen zur Disposition gestellt.

Tab. 4.10: Korrelationen mit dem Konstrukt „Interesse“

ML-Methode		ADF-Methode	
Konstrukt	Korrelation	Konstrukt	Korrelation
Interessantheit	0,691	Interessantheit	[0,780]
Verknüpfung	0,570	Verknüpfung	[0,682]
Materialien	0,544	Unterrichtsführung	[0,665]
Unterrichtsführung	0,534	Materialien	[0,662]
Feedback	0,520	Sozialklima	[0,606]
Sozialklima	0,506	Feedback	[0,593]
-----		-----	
		Fachkompetenz	[0,590]
		Schülerautonomie	[0,580]
		Diagnostik	[0,555]
		Selbstengagement	[0,552]
		Zeitmanagement	[0,536]
		Leistungserwartungen	[-0,509]

In Tabelle 4.10 sind alle Korrelationen über 0,5, d.h. alle mindestens als mittlere Korrelationen einzuschätzenden Werte (Bühl 2006, S. 342), mit dem Konstrukt „Interesse“ in Abhängigkeit der Schätzmethode dargestellt. Bei der ML-Methode werden entsprechende Korrelationen bei den Konstrukten „Interessantheit“, „Verknüpfung“, „Materialien“, „Unterrichtsführung“, „Feedback“ und „Sozialklima“ geschätzt und damit kann davon ausgegangen werden, dass auch zwischen den Konstrukten aus den Konstruktgruppen „Lehrerpersönlichkeit“ und „Unterrichtsangebot“ zahlreiche Zusammenhänge bestehen. Die gleichen sechs Konstrukte besitzen auch bei der ADF-Methode die höchsten Korrelationen, wobei die Höhe unterschiedlich ist und außer-

dem noch Korrelationen über 0,5 für die sechs Konstrukte „Fachkompetenz“, „Schülerautonomie“, „Diagnostik“, „Selbstengagement“, „Zeitmanagement“ und „Leistungserwartungen“ geschätzt werden.

Werden nun zur Beseitigung von Multikollinearität bei Verwendung der ML-Methode die Konstrukte mit Korrelationen über 0,5 mit dem Konstrukt „Interesse“ entfernt, wobei die Konstrukte entsprechend der Höhe der Korrelationen sukzessive aus dem Modell entfernt werden, d.h. nach und nach wird das Konstrukt mit der höchsten Korrelation entfernt, bestehen in den jeweiligen Modellen weiterhin nur Korrelationen über 0,5 bei den in Tabelle 4.10 aufgeführten Konstrukten im Bezug auf das Interesse. Aus diesem Grund werden alle sechs Konstrukte komplett aus dem Modell entfernt. Bei Verwendung der ADF-Methode wurde analog vorgegangen, allerdings zeigte sich, dass nach der Entfernung der sechs Konstrukte mit den höchsten Korrelationen die Korrelationen der Konstrukte „Fachkompetenz“, „Schülerautonomie“, „Diagnostik“, „Selbstengagement“, „Zeitmanagement“ und „Leistungserwartungen“ mit dem Konstrukt „Interesse“ nicht mehr über 0,5 geschätzt werden. Aus diesem Grund werden diese Konstrukte auch zum jetzigen Zeitpunkt nicht aus dem Modell entfernt. Das neue Strukturmodell (2), welches auf Basis der Ergebnisse der ML- und ADF-Methode identisch modifiziert wurde, wird nun wieder im Hinblick auf die Zuverlässigkeit der Schätzungen und die Modellgüte untersucht.

Betrachtet man nun in Tabelle 4.11 die Ergebnisse der ML-Methode bei dem modifizierten Strukturmodell lässt sich feststellen, dass die Modellgüte tendenziell gleich geblieben ist und die Reliabilität der endogenen latenten Variablen von 0,800 im Ausgangsmodell auf 0,740 gesunken ist. Weiterhin ist bei den Variablen „Schüleraktivierung“ und „Binnendifferenzierung“ wie im Vorherigen und nun bei der Variablen „Zeitmanagement“ nicht auszuschließen, dass die geschätzten Regressionskoeffizienten gleich null sind. Im SGM 2 ist auch nur noch die Richtung der Beziehung mit dem Lernzuwachs des Konstrukts „Leistungserwartungen“ nicht der am Anfang getroffenen Hypothese entsprechend. Die Modellgüte beim modifizierten Modell bei Verwendung der ADF-Methode hat sich im Vergleich zum Ausgangsmodell deutlich

verbessert, wobei die Variation der endogenen Variable nun um rund 10% schlechter durch die einfließenden exogenen Variablen erklärt werden kann. Auch können durch die angesprochenen Modifikationen des Modells mehr der anfänglich unterstellten Richtungen der Beziehungen zwischen der endogenen und den exogenen Variablen bestätigt werden. Da aber auf Basis der drei Kriterien (SRMR, CFI und TLI) zur Beurteilung der Modellgüte und residualen Kovarianzen über 0,1 nicht von einer akzeptablen Modellgüte bzw. -anpassung ausgegangen werden kann sowie weiterhin einige Richtungen der Beziehungen nicht plausibel sind, wird das zweite Modell weiter modifiziert.

Tab. 4.11: Ergebnisse der Parameterschätzung des SGM bei Verwendung der ML-Methode [ADF-Methode] – Modell 2

exogenes Konstrukt	C.R. der geschätzten Parameter		standardisiertes Regressionsgewicht	
Fachkompetenz	15,211	[7,345]	0,210	[0,126]
Diagnostik	15,565	[11,582]	0,191	[0,148]
Schüleraktivierung	0,347	[0,186]	0,004	[0,002]
Schülerautonomie	2,917	[3,461]	0,043	[0,054]
Mindestanforderungen	-9,103	[-0,567]	-0,084	[-0,008]
Leistungserwartungen	3,447	[1,966]	0,040	[0,029]
Binnendifferenzierung	0,929	[2,460]	0,008	[0,021]
Zeitmanagement	0,802	[2,409]	0,020	[0,033]
Selbstengagement	7,469	[-0,038]	0,070	[-0,001]
Leistungsfähigkeit	10,276	[10,631]	0,125	[0,148]
Interesse	46,194	[40,591]	0,446	[0,600]
Schulbeurteilung	8,693	[2,652]	0,058	[0,020]
Reliabilität der endogenen latenten Variable: 0,740 [0,747]				
SRMR: 0,051 [0,052] CFI: 0,899 [0,726] TLI: 0,833 0,682]				

Im Rahmen einer erneuten Modifizierung des SGM werden nun sukzessive die höchsten Korrelationen über 0,5 identifiziert, dann jeweils das exogene Konstrukte aus dem Messmodell ausgeschlossen, dass das niedrigere standardisierte Regressionsgewicht der beiden entsprechenden Konstrukte im Modell besitzt, und das modifizierte SGM erneut geprüft. In Tabelle 4.12 sind die Korrelationen über 0,5 dargestellt. Als erstes wurde die Korrelation zwischen

den Konstrukten „Fachkompetenz“ und „Zeitmanagement“ identifiziert. Diese Korrelation ist sowohl bei der ML- als auch bei der ADF-Methode die höchste zwischen den exogenen Konstrukten. Da das markierte Konstrukt „Zeitmanagement“ ein niedrigeres Regressionsgewicht besitzt, wurde es als erstes aus dem zuletzt angenommenen Modell entfernt. Anschließend wurden die in Tabelle 4.12 dargestellten Korrelationen in absteigender Reihenfolge identifiziert und es wurde analog zu der angesprochenen Ausschlussmethodik verfahren. Die schon im Vorherigen angesprochenen hohen Abhängigkeiten vor allem zwischen den Konstrukten aus den Konstruktgruppen „Lehrerpersönlichkeit“ und „Unterrichtsangebot“ können auch durch die in Tabelle 4.12 aufgeführten Korrelationen bestätigt werden. Auch zeigte sich eine deutliche Abhängigkeit der Konstrukte „Fachinteresse“ und „Selbstengagement“, was zu einem Ausschluss der Variablen „Selbstengagement“ führte. Dieser Ausschluss kann damit begründet werden, dass das Interesse bezüglich eines Faches das eigene Engagement im Bezug auf dieses Fach logischer Weise positiv beeinflussen sollte. Mit der Korrelation zwischen diesen beiden Konstrukten zeigte sich außerdem, dass auch zwischen den Variablen aus den Konstruktgruppen „Unterrichtsnutzung“ und „Schülervoraussetzungen“ deutliche Abhängigkeiten bestehen.

Tab. 4.12: Korrelationen der exogenen latenten Variablen bei Verwendung der ML-Methode [ADF-Methode] – Modell 2

Konstrukte		Korrelationen	
Fachkompetenz	Zeitmanagement	0,710	[0,665]
Schülerautonomie	Schüleraktivierung	0,658	[0,659]
Diagnostik	Schülerautonomie	0,616	[0,561]
Leistungsfähigkeit	Leistungserwartungen	-0,587	[-0,585]
Fachkompetenz	Diagnostik	0,567	[0,537]
Fachkompetenz	Mindestanforderungen	-0,501	[-0,509]
Interesse	Selbstengagement	0,464	[0,529]

Im Zuge der Entfernung der Variablen „Zeitmanagement“, „Schüleraktivierung“, „Schülerautonomie“, „Leistungserwartungen“, „Diagnostik“, „Mindestanforderungen“ und „Selbstengagement“ aus dem Messmodell konnte die Mo-

dellgüte auf Basis des CFI und TLI bei beiden Methoden deutlich im Vergleich zum Modell 2 verbessert werden (siehe Tabelle 4.13). Nur bei Verwendung der ML-Methode musste dadurch eine leichte Abnahme der Reliabilität im Bezug auf den Lernzuwachs hingenommen werden. Weiterhin liefern bei beiden Methoden auf Basis der C.R. der geschätzten Regressionskoeffizienten die einzelnen exogenen latenten Variablen signifikante Beiträge zur Erklärung der Variation des Lernzuwachses, auch die anfänglich ausgestellten Hypothesen bezüglich der Richtungen der Zusammenhänge zwischen der endogenen und den exogenen Variablen werden nun alle durch das Modell bestätigt und es sind keine Korrelationen größer als 0,5 mehr zwischen den exogenen latenten Variablen geschätzt worden. Zwar kann auf Basis der Fit-Indizes CFI und TLI nicht von einer insgesamt akzeptablen Modellgüte ausgegangen werden, da sich für die Indizes Werte unter 0,95 ergeben haben, aber die residualen Kovarianzen sprechen auf Grund von Werten überwiegend deutlich unter 0,05 für ein gutes Modell bzw. eine gute Modellanpassung und es werden daher keine Modifikationen mehr vorgenommen.

Tab. 4.13: Ergebnisse der Parameterschätzung des SGM bei Verwendung der ML-Methode [ADF-Methode] – Modell 3

exogenes Konstrukt	C.R. der geschätzten Parameter		standardisiertes Regressionsgewicht	
Fachkompetenz	40,020	[23,741]	0,355	[0,276]
Binnendifferenzierung	7,937	[9,549]	0,057	[0,074]
Leistungsfähigkeit	16,111	[15,282]	0,130	[0,130]
Interesse	56,468	[51,954]	0,529	[0,627]
Schulbeurteilung	13,726	[4,173]	0,094	[0,031]
Reliabilität der endogenen latenten Variable: 0,695 [0,747]				
SRMR: 0,048 [0,055] CFI: 0,936 [0,860] TLI: 0,920 [0,825]				

Mit Hilfe des Modells 3 mit den exogenen latenten Variablen „Fachkompetenz“, „Binnendifferenzierung“, „Leistungsfähigkeit“, „Interesse“ sowie „Schulbeurteilung“ kann rund 70% der Variation der endogenen latenten Variable „Lernzuwachs“ erklärt werden. Den mit Abstand größten Einfluss auf den wahrgenommenen Lernzuwachs von Schülerinnen und Schülern im Unterricht

ihres Lehrers hat das „fachbezogenes *Interesse*“ der Schülerinnen und Schüler. Die Fachkompetenz des Lehrers beeinflusst den Lernzuwachs am zweit meisten und die selbsteingeschätzte Leistungsfähigkeit der Schülerinnen und Schüler am dritt meisten. Auch die Konstrukte „Binnendifferenzierung“ und „Einstellung zur Schule“ (*Schulbeurteilung*)“ haben einen signifikanten, wenn auch relativ geringen, Einfluss. Durch die in Tabelle 4.10 und 4.12 aufgeführten Korrelationen zwischen den exogenen Variablen muss jedoch gefolgert werden, dass die Konstrukte „fachbezogenes *Interesse*“ und „Fachkompetenz“ nicht nur für diese Aspekte der Unterrichtsqualität stehen, sondern für die Persönlichkeit der Lehrerinnen und Lehrer und deren Unterrichtsweise insgesamt. So ist beim Konstrukt „Interesse“ auf Grund der Korrelationen davon auszugehen, dass die Schülerinnen und Schüler weniger in Abhängigkeit ihres schon vorhandenen Interesses antworteten, sondern vielmehr in Abhängigkeit ihrer Interesses infolge eines beispielsweise spannenden und anregenden Unterrichts. Und für das Konstrukt „Fachkompetenz“ muss unterstellt werden, dass die fachliche Kompetenz einer Lehrerin bzw. eines Lehrers auch in Abhängigkeit der Fähigkeit Inhalte zu vermitteln von den Schülerinnen und Schülern gesehen wird. Damit zeigte sich analog zu den Ausführungen von Wagner (2008), dass die einzelnen Merkmale der Unterrichtsqualität, die aus der Sicht der Lehrerinnen und Lehrer wichtig zur Erfassung der Qualität von Unterricht sind und zur Bildung der 19 unterstellten Konstrukte führten, von den Schülerinnen und Schülern nicht in der erhofften Weise unterschieden werden können. Zur weiteren Überprüfung dieser Aussagen könnte es ggf. helfen ein Strukturgleichungsmodell aufzustellen, bei dem alle Konstrukte mit Bezug zur Persönlichkeit der Lehrerinnen und Lehrer und deren Unterrichtsweise zu einem übergeordneten Konstrukt zusammenfasst werden und das übergeordnete Konstrukt dann zur Erklärung der Variation beim Konstrukt „Lernzuwachs“ genutzt wird. Zusätzlich könnten somit die Probleme von Multikollinearität umgangen werden.

Abschließend werden nun Unterschiede zwischen den Ausprägungen der Variablen „Geschlecht“, „Bildungsgang“, „Klassenstärke“ und „Notenschnitt“ (in Kategorien) im Bezug auf den wahrgenommenen Lernzuwachs mit Hil-

fe von SGM untersucht. Dazu wurden je Variable entsprechend der Anzahl der Merkmalsausprägungen (siehe Abschnitt 2.2) Dummy-Variablen erzeugt, wobei die Kategorie „Unbekannt“ durch die Dummy-Variablen der anderen Merkmalsausprägungen bestimmt wird. Die Modelle konnten nur mit der ML-Methode geschätzt werden, da die Gewichtungsmatrix der ADF-Schätzmethode bei diesen Strukturgleichungsmodellen nicht invertiert werden konnte. Einerseits zeigen die in Tabelle 4.14 dargestellten Ergebnisse, dass alle hier untersuchten SGM eine sehr gute Modellgüte besitzen. Andererseits kann durch die Modelle fast keine Variation bei der latenten Variablen „Lernzuwachs“ erklärt werden. Nur bei dem SGM mit den Dummy-Variablen des Geschlechts ist die Dummy-Variable „männlich – ja/nein“ und bei dem SGM mit den Dummy-Variablen des Notenschnitts ist die Dummy-Variable „Notenschnitt 3 – ja/nein“ nicht signifikant unterschiedlich von null. Weiterhin sind nur im SGM der Variablen „Notenschnitt“ die Regressionsgewichte der Dummy-Variablen „Notenschnitt 4 – ja/nein“, „Notenschnitt 5 – ja/nein“ sowie „Notenschnitt 5 – ja/nein“ negativ und die restlichen Regressionsgewichte aller SGM sind entweder wie bereits erwähnt null oder positiv. Damit kann auf Basis dieser Erkenntnisse davon gesprochen werden, dass das Geschlecht, der Bildungsgang, die Klassenstärke und der Notenschnitt keine Einflüsse auf den wahrgenommene Lernzuwachs haben, obwohl Unterschiede in den Wahrnehmung der Schülerinnen und Schüler bestehen.

Tab. 4.14: Ergebnisse der Parameterschätzung der Dummy-SGM bei Verwendung der ML-Methode

Variable	SRMR	CFI	TLI	Reliabilität
Geschlecht	0,007	1,000	0,999	0,004
Bildungsgang	0,008	0,995	0,984	0,027
Klassenstärke	0,007	0,996	0,991	0,034
Notenschnitt	0,008	0,994	0,991	0,019

5 Zusammenfassung

Mit Hilfe von verschiedenen statistischen Verfahren wurden im Rahmen dieser Arbeit die Evaluationsdaten von 23 Berliner Schulen ausgewertet. Dabei wurde zunächst der Fragestellung nachgegangen, ob anhand der gewählten Befragungsinhalte Wahrnehmungen von Schülerinnen und Schülern im Bezug auf verschiedene, nicht direkt messbare Aspekte der Unterrichtsqualität erfasst werden können. Darauf aufbauend wurden mögliche Einflussfaktoren auf den Erfolg von Schule gemessen durch den wahrgenommenen Lernzuwachs untersucht.

Die erste Fragestellung wurde mit Hilfe von konfirmatorischen Faktorenanalysen unter Berücksichtigung verschieden Schätzverfahren untersucht. Dabei wurde zunächst überprüft, ob die auf Basis von sachlogischen Überlegungen im Fragebogen enthaltenen Items zur Abbildung der im Vorfeld bestimmten Konstrukte geeignet sind. Da dabei nur für Teile der Konstrukte die Messbarkeit anhand aller entsprechenden Items empirisch bestätigt werden konnte, wurden die gewonnenen Erkenntnisse zur Modifikation der Modelle zur Messung der Konstrukte verwendet. Letztendlich konnten anhand der Ergebnisse der konfirmatorischen Faktorenanalysen für die überwiegende Zahl der Konstrukte eine adäquate Messung unterstellt werden. Problematisch war jedoch das einige Konstrukte nur durch drei Items indirekt gemessen wurden und damit die unterstellten Modelle nicht abschließend mit Hilfe von globalen Gütekriterien beurteilt werden konnten.

Die Analyse der Abhängigkeiten auf Basis der Konstrukte insbesondere im Bezug auf den Erfolg von Schule gemessen durch den wahrgenommenen Lernzuwachs wurde mit Strukturgleichungsmodellen unter Berücksichtigung verschiedener Schätzverfahren durchgeführt. Letztendlich konnten als Hauptein-

flussfaktoren auf den wahrgenommenen Lernzuwachs die Konstrukte „Fachbezogenes Interesse“ und „Fachkompetenz“ identifiziert werden, wobei diese Konstrukte als Indikatoren für die Persönlichkeit der Lehrerinnen und Lehrer und deren Unterrichtsweisen insgesamt stehen. Außerdem konnte nachgewiesen werden, dass beim Geschlecht sowie Bildungsgang, der Klassenstärke und dem Notenschnitt Unterschiede im Bezug auf die Beurteilung des Lernzuwachses bestehen.

Basierend auf allen in dieser Arbeit durchgeführten Analysen eignet sich die Befragung insbesondere zur Abbildung der Wahrnehmung der Schülerinnen und Schüler bezüglich der empfundenen Qualität des Unterrichts in einer Klasse, wenn damit den jeweiligen Lehrerinnen und Lehrern ein Feedback auf Klassenebene gegeben werden soll. Damit wird das originäre Ziel der Befragung erreicht. Zur Analyse der Einflussfaktoren auf den wahrgenommenen Lernzuwachs der Schülerinnen und Schüler ist diese Befragung allerdings weniger geeignet, da die Schülerinnen und Schüler den Lernzuwachs hauptsächlich in Abhängigkeit der Persönlichkeit ihrer Lehrerin oder ihres Lehrers wahrnehmen und weniger anhand der einzelnen Merkmale von Unterrichtsqualität.

Literaturverzeichnis

- [1] ARBUCKLE, J. L. (2007): *Amos 16.0 User's Guide*. AMOS Development Corporation, www.amosdevelopment.com/download/Amos%2016.0%20User%27s%20Guide.pdf.
- [2] BACKHAUS, K./ERICHSON, B./PLINKE, W./WEIBER, R. (2006): *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 11. Aufl., Springer.
- [3] BACKHAUS, K./ERICHSON, B./PLINKE, W./WEIBER, R. (2008A): *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 12. vollst. überarb. Aufl., Springer, Online Kapitel 12, www.multivariate.de.
- [4] BACKHAUS, K./ERICHSON, B./PLINKE, W./WEIBER, R. (2008B): *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 12. vollst. überarb. Aufl., Springer, Online Kapitel 11, www.multivariate.de.
- [5] BERLINER SENATSVERWALTUNG FÜR BILDUNG, WISSENSCHAFT UND FORSCHUNG (HRSG., 2009) : *Schulgesetz für das Land Berlin*. vom 26. Januar 2004 (GVBl. S. 26), zuletzt geändert durch Artikel I des Gesetzes am 2. März 2009 (GVBl. S. 62), www.berlin.de/imperia/md/content/sen-bildung/rechtsvorschriften/schulgesetz.pdf.
- [6] BÜHL, A.: *SPSS 14 – Einführung in die moderne Datenanalyse*. 10. überarb. und erw. Aufl., Pearson Studium, München.

- [7] BÜHNER, M. (2006): *Einführung in die Test- und Fragebogenkonstruktion*. 2. akt. Aufl., Pearson Studium, München.
- [8] BOLLEN, K. A./STINE, R. A. (1993): *Bootstrapping Goodness-of-Fit Measure in Structural Equation Modeling*. in: Bollen, K. A./ Long, J. S. (Hrsg.), *Testing Structural Equations Models*, Newbury Park, S. 111-135.
- [9] BOOMSMA, A. (1982): *The Robustness of LISREL against Small Sample Size in Factor Analysis Models*. in: Jöreskog, K. G./ Wold, H. (Hrsg.), *Systems Under Indirect Observations: Causality, Structure, Prediction – Part I*, Amsterdam, S. 149-174.
- [10] BROWNE, M. W. (1984): *Asymptotically distribution-free Methods for the Analysis of Covariance Structures*. *British Journal of Mathematical and Statistical Psychology*, Jg. 37, S. 62-83.
- [11] CURRAN, P. J./WEST, S. G./FINCH, J. F. (1996): *The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis*. *Psychological Methods*, Jg. 1, Nr. 1, S. 16-29.
- [12] DECARLO, L. T. (1997): *On the Meaning and Use of Kurtosis*. *Psychological Methods*, Jg. 2, Nr. 3, S. 292-307.
- [13] DITTON, H./MERZ, D. (2001): *Qualitätssicherung in Schule und Unterricht – Fragebogen für Schülerinnen und Schüler – Fragen zum Unterricht*. Universität Osnabrück.
- [14] GREWAL, R./COTE, J.A./BAUMGARTNER, H. (2004): *Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing*. *Marketing Science*, Jg. 23, Nr. 4, S. 519-529.
- [15] HARTUNG, J./ELPELT, B./KLÖSENER, K.-H. (2005): *Statistik – Lehr- und Handbuch der angewandten Statistik*. Oldenbourg.
- [16] HELMKE, A. (2003): *Unterrichtsqualität erfassen, bewerten, verbessern*. Kallmeyer Verlagsbuchhandlung, Seelze-Velber.

- [17] HU, L./BENTLER, P.M. (1998): *Fit Indices in Covariance Structure Analysis: Sensitivity to Underparameterized Model Misspecification*. Psychological Methods, Jg. 3, Nr. 4, S. 424-453.
- [18] HU, L./BENTLER, P.M. (1999): *Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives*. Structural Equation Modeling: A Multidisciplinary Journal, Jg. 6, Nr. 1, S. 1-55.
- [19] MARDIA, K. V. (1970): *Measures of Multivariate Skewness and Kurtosis with Applications*. Biometrika, Jg. 57, Nr. 3, S. 519-530.
- [20] MARDIA, K. V. (1974): *Applications of some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies*. Sankhyā: The Indian Journal of Statistics, Jg. 36B, Nr. 2, S. 115-128.
- [21] MARUYAMA, G. M. (1998): *Basics of Structural Equation Modeling*. SAGE, Thousand Oaks, CA.
- [22] NEVITT, J./HANCOCK, G. R. (2001): *Performance of Bootstrapping Approaches to Model Test Statistics and Parameter Standard Error Estimation in Structural Equation Modeling*. Structural Equation Modeling: A Multidisciplinary Journal, Jg. 8, Nr. 3, S. 353-377.
- [23] OLSSON, U. H./FOSS, T./TROYE, S. V./HOWELL, R.D. (2000): *The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality*. Structural Equation Modeling: A Multidisciplinary Journal, Jg. 7, Nr. 4, S. 557-595.
- [24] OSZ GESUNDHEIT I (2008): *Evaluationsbericht*. Berlin, www.osz-gesundheit.de/organisation/schulprogramm/Evaluationsbericht%202008%202009.pdf.
- [25] SCHERMELLEH-ENGEL, K./MOOSBRUGGER, H./MÜLLER, H. (2003): *Evaluating the Fit of Structural Equations Models: Test*

- of Significance and Descriptive Goodness-of-Fit Measures.* Methods of Psychological Research Online, Jg. 8, Nr. 2, S. 23-74.
- [26] SEEBER, S./SQUARRA, D. (2003): *Lehren und Lernen in beruflichen Schulen, Schülerurteile zur Unterrichtsqualität.* Lang, Frankfurt.
- [27] THURSTONE, L. L. (1931): *Multiple Factor Analysis.* Psychological Review, Jg. 38, S. 406-427.
- [28] WAGNER, C. (2008): *Netzwerk Schülerbefragung: Schülerurteile zur Unterrichtsqualität – Vorläufiger Bericht der Gesamtauswertung zur Befragung im Dezember 2008.* Humboldt-Universität zu Berlin (Unveröffentlichter Bericht).
- [29] WAGNER, C. (2009A): *Unterrichtsentwicklung durch Schülerbefragungen? Rezeption von Ergebnissen aus der internen Evaluation im Rahmen von Schulentwicklung.* in: van Buer, J./Rückmann, J. (Hrsg.), *Pädagogische Schulentwicklung beruflicher Schulen in Berlin – Steuerung, Selbstregulation, organisatorisches Lernen im Modellprojekt 'Qualitätsentwicklung beruflicher Schulen' (QEBS), Studien zur Wirtschaftspädagogik und Berufsbildungsforschung aus der Humboldt-Universität zu Berlin, Bd. 15, Berlin, S. 157-174.*
- [30] WAGNER, C. (2009B): *Persönliche Kommunikation.* Berlin (Wagner, C.: Wissenschaftliche Mitarbeiterin am Institut für Erziehungswissenschaften der Humboldt-Universität zu Berlin).

Anhang A – Schülerfragebogen

Netzwerk Berliner Schulen für die interne Evaluation		Schülerinnen-/Schülerbefragung			
	Inwieweit treffen die folgenden Aussagen auf Ihrer Lehrerin/Ihren Lehrer zu? Meine Lehrerin/mein Lehrer ...	Trifft zu (4)	Trifft eher zu (3)	Trifft eher nicht zu (2)	Trifft nicht zu (1)
le04	- nimmt Beiträge und Ideen von uns Schülern/Schülerinnen in den Unterricht auf.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ug03	- weiß genau, bei welchen Aufgaben ich Schwierigkeiten habe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uh04	- ist gerecht zu uns, sodass wir ihm/ihr vertrauen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uh03	- nimmt sich Zeit, wenn man etwas mit ihm/ihr besprechen möchte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ug01	- weiß genau, was ich leisten kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
le01	- und die Klasse diskutieren gemeinsam.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ui03	- kann auf jede fachliche Frage eine Antwort geben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ub05	- erklärt uns, warum der behandelte Inhalt wichtig ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ub02	- benutzt Bilder und Beispiel, um den Lehrstoff zu veranschaulichen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ud04	- kann sich nicht in der Klasse durchsetzen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ua05	- informiert uns über den geplanten Unterrichtsverlauf.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uf01	- stellt im Unterricht geringe Anforderungen, sodass ich mich langweile.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uf03	- stellt so einfache Aufgaben, dass ich gar nicht wirklich nachdenken muss.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
le02	- ermöglicht es uns, eigene Vorschläge in die Unterrichtsgestaltung einzubringen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ua01	- spricht laut und deutlich, sodass ich alles gut verstehe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uf02	- geht im Stoff so langsam voran, dass ich keine Lust habe, dem Unterricht zu folgen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ua04	- geht im Unterricht in einer logischen Reihenfolge vor.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ui01	- ist immer gut auf den Unterricht vorbereitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ua02	- drückt sich klar und verständlich aus.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
le03	- lässt uns im Unterricht Fragen stellen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ub03	- erklärt uns, wofür wir die behandelten Inhalte verwenden können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uh02	- ermöglicht es uns, über Probleme zu reden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ue03	- lässt uns schwierige Klassenarbeiten/Tests schreiben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ui02	- beherrscht den Lehrstoff.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ud05	- legt Wert auf einen pünktlichen Beginn des Unterrichts.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ud01	- hat die Klasse im Griff.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ua06	- unterteilt den Unterricht in überschaubare Abschnitte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ue01	- stellt oft so schwere Aufgaben, dass ich sie nicht lösen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
le05	- lässt uns einzelne Unterrichtsstunden mitgestalten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ud02	- schafft es, für Ruhe und Ordnung zu sorgen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ug02	- bemerkt Veränderungen in meinen Leistungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uc04	- gestaltet den Unterricht interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ue02	- gestaltet den Unterricht so schwer, dass ich nicht mitkomme.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ub04	- stellt Beziehungen zwischen den behandelten Themen heraus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uc02	- stellt uns interessante Aufgaben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ua03	- gibt uns klare und verständliche Arbeitsanweisungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ua07	- gibt uns einen Überblick zur Gliederung des Stoffs.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uc01	- gestaltet den Unterricht abwechslungsreich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ub01	- stellt Verbindungen zwischen dem Unterrichtsstoff und dem Berufs- bzw. Alltagsleben her.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ud03	- fordert, dass sich alle Schüler/Schülerinnen aktiv am Unterricht beteiligen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uh01	- bemerkt, wenn es Probleme in der Klasse gibt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uc03	- stellt Bezüge zu aktuellen Themen her.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Inwieweit treffen die folgenden Aussagen auf den Unterricht Ihrer Lehrerin/Ihren Lehrer zu?	Trifft zu	Trifft eher zu	Trifft eher nicht zu	Trifft nicht zu
if04	Ich habe im Unterricht die Möglichkeit, an selbst gewählten Aufgaben zu arbeiten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
if03	Bei umfangreichen Aufgaben kann ich mir die Arbeit selbst einteilen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ld02	Wir führen eigene Untersuchungen/Projekte durch.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ld01	Wir bearbeiten Aufgaben in Gruppen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lb03	Arbeitsgruppen werden nach dem Können der Schüler/-innen zusammen gestellt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lb01	Bessere Schüler erhalten schwierigere Aufgaben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lc03	Jeder arbeitet für sich an den gleichen Aufgaben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
la03	Das Unterrichtsmaterial hilft mir, den Lehrstoff der Stunde zu wiederholen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
if05	Wir beurteilen uns gegenseitig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lb02	Wir arbeiten je nach unserem Können an unterschiedlich schweren Aufgaben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
la02	Das Unterrichtsmaterial ist übersichtlich gegliedert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lc01	Der Lehrer/die Lehrerin redet und stellt Fragen, einzelne Schüler/-innen antworten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ld03	Aus dem Unterricht meines Lehrers/meiner Lehrerin kenne ich eine oder mehrere der folgenden Methoden: Fishbowl, Stamm-Experten-Gruppe, Mind Mapping, Stationen Lernen, schnelle Lesemethode, Doppelkreis.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
if02	Die Gesprächsleitung haben manchmal die Schüler.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ld04	Wir bearbeiten Aufgaben mit Hilfe des Computers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
if01	Der Lehrer/die Lehrerin lässt uns eigene Lösungswege finden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lc02	Wir sitzen und hören zu, der Lehrer/die Lehrerin redet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ld05	Mein Lehrer/meine Lehrerin lässt uns unsere Arbeitsergebnisse vorstellen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
la01	Mein Lehrer/meine Lehrerin stellt uns Unterrichtsmaterialien zu allen wichtigen Themen zur Verfügung.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lc04	Wir bearbeiten mit Hilfe des Lehrers/der Lehrerin Arbeitsblätter.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
la04	Das Unterrichtsmaterial hilft mir, Klausuren vorzubereiten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Inwieweit treffen die folgenden Aussagen auf Sie zu?				
mb02	In Klassenarbeiten schreibe ich häufig gute Noten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
za04	Ich glaube, dass ich das, was ich im Unterricht gelernt habe, später gut verwenden kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
za01	Wenn ich auf das letzte Halbjahr zurück blicke, habe ich viel dazu gelernt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mb01	Ich gehöre zu den leistungsstarken Schülern/Schülerinnen in meiner Klasse.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ga01	Mit meiner Schule bin ich sehr zufrieden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
md02	Im Unterricht in diesem Fach frage ich mich oft, was ich hier soll.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mb03	Lernen fällt mir leicht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
md03	Über die Inhalte dieses Faches würde ich gerne noch mehr erfahren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mc04	Auf Klausuren/Tests bereite ich mich in der Regel gut vor.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ma04	Der Lehrer/die Lehrerin gibt ausreichend Rückmeldung über meinen Leistungsstand.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
md04	Die Inhalte in diesem Fach finde ich interessant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mc01	Meine Hausaufgaben erledige ich in der Regel sorgfältig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mb04	Ich brauche immer viel Zeit, um Lerninhalte zu verstehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
md01	Die in diesem Unterrichtsfach behandelten Inhalte machen mir Spaß.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ma01	Ich verstehe, wie meine Noten zustande kommen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
za03	Im Unterricht habe ich gelernt, die verschiedenen Fachgebiete zu verknüpfen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ma02	Ich weiß, welche Leistungen im Unterricht von mir erwartet werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mc03	Im Unterricht arbeite ich meist konzentriert mit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ga03	Wenn mein bester Freund/meine beste Freundin fragt, würde ich diese Schule empfehlen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ma03	Ich kenne meine Stärken und Schwächen im Unterricht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
za02	Bei diesem Lehrer/bei dieser Lehrerin habe ich wirklich viel gelernt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ga02	Wenn ich die Wahl hätte, würde ich wieder auf diese Schule gehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mc02	Im Unterricht lasse ich mich leicht ablenken.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
geschl	Ich bin	<input type="checkbox"/> männlich <input type="checkbox"/> weiblich			
note	Meine Schulnoten lagen im letzten Zeugnis im Durchschnitt bei (z.B. 3,5) ...	<input type="text"/> , <input type="text"/>			

Anhang B – Detaillierte Item-Konstrukt-Zuordnung

Die im Rahmen dieser Arbeit unterstellte Item-Konstrukt-Zuordnung der Schülerbefragung des Schuljahres 2007/2008 ist folgende:

Lehrerpersönlichkeit:

- Fachkompetenz
 - ui01: Meine Lehrerin/mein Lehrer ist immer gut auf den Unterricht vorbereitet.
 - ui02: Meine Lehrerin/mein Lehrer beherrscht den Lehrstoff.
 - ui03: Meine Lehrerin/mein Lehrer kann auf jede fachliche Frage eine Antwort geben.
- Unterrichtsführung und Strukturiertheit
 - ua01: Meine Lehrerin/mein Lehrer spricht laut und deutlich, sodass ich alles gut verstehe.
 - ua02: Meine Lehrerin/mein Lehrer drückt sich klar und verständlich aus.
 - ua03: Meine Lehrerin/mein Lehrer gibt uns klare und verständliche Anweisungen.
 - ua04: Meine Lehrerin/mein Lehrer geht im Unterricht in einer logischen Reihenfolge vor.

- ua05: Meine Lehrerin/mein Lehrer informiert uns über den geplanten Unterrichtsverlauf.
- ua06: Meine Lehrerin/mein Lehrer unterteilt den Unterricht in überschaubare Abschnitte.
- ua07: Meine Lehrerin/mein Lehrer gibt uns einen Überblick zur Gliederung des Stoffs.
- Diagnostische Kompetenz (Diagnostik)
 - ug01: Meine Lehrerin/mein Lehrer weiß genau, was ich leisten kann.
 - ug02: Meine Lehrerin/mein Lehrer bemerkt Veränderungen in meinen Leistungen.
 - ug03: Meine Lehrerin/mein Lehrer weiß genau, bei welchen Aufgaben ich Schwierigkeiten habe.
- Lehrer-Schüler-Sozialklima
 - uh01: Meine Lehrerin/mein Lehrer bemerkt, wenn es Probleme in der Klasse gibt.
 - uh02: Meine Lehrerin/mein Lehrer ermöglicht uns, über Probleme zu reden.
 - uh03: Meine Lehrerin/mein Lehrer nimmt sich Zeit, wenn man etwas mit ihm/ihr besprechen möchte.
 - uh04: Meine Lehrerin/mein Lehrer ist gerecht zu uns, sodass wir ihm/ihr vertrauen.

Unterrichtsangebot:

- Veranschaulichung
 - ub01: Meine Lehrerin/mein Lehrer - stellt Verbindungen zwischen dem Unterrichtsstoff und dem Berufs- bzw. Alltagsleben her.

- ub02: Meine Lehrerin/mein Lehrer benutzt Bilder und Beispiel, um den Lehrstoff zu veranschaulichen.
 - ub03: Meine Lehrerin/mein Lehrer erklärt uns, wofür wir die behandelten Inhalte verwenden können.
 - ub04: Meine Lehrerin/mein Lehrer stellt Beziehungen zwischen den behandelten Themen heraus.
 - ub05: Meine Lehrerin/mein Lehrer erklärt uns, warum behandelte Inhalt wichtig ist.
- Interessantheit
 - uc01: Meine Lehrerin/mein Lehrer gestaltet den Unterricht abwechslungsreich.
 - uc02: Meine Lehrerin/mein Lehrer stellt uns interessante Aufgaben.
 - uc03: Meine Lehrerin/mein Lehrer stellt Bezüge zu aktuellen Themen her.
 - uc04: Meine Lehrerin/mein Lehrer gestaltet den Unterricht interessant.
- Transparenz und Feedback
 - ma01: Ich verstehe, wie meine Noten zustande kommen.
 - ma02: Ich weiß, welche Leistungen im Unterricht von mir erwartet werden.
 - ma03: Ich kenne meine Stärken und Schwächen im Unterricht.
 - ma04: Der Lehrer/die Lehrerin gibt ausreichend Rückmeldung über meinen Leistungsstand.
- Unterstützende Materialien
 - la01: Meine Lehrerin/mein Lehrer stellt uns Unterrichtsmaterialien zu allen wichtigen Themen zur Verfügung.

- la02: Das Unterrichtsmaterial ist übersichtlich gegliedert.
- la03: Das Unterrichtsmaterial hilft mir, den Lehrstoff der Stunde zu wiederholen.
- la04: Das Unterrichtsmaterial hilft mir, Klausuren vorzubereiten.
- Schüleraktivierender Unterricht (Schüleraktivierung)
 - ld01: Wir bearbeiten Aufgaben in Gruppen.
 - ld02: Wir führen eigene Untersuchungen/Projekte durch.
 - ld03: Aus dem Unterricht meines Lehrers/meiner Lehrerin kenne ich eine oder mehrere der folgenden Methoden: Fishbowl, Stamm-Experten-Gruppe, Mind Mapping, Stationen Lernen, schnelle Lesemethode, Doppelkreis.
 - ld05: Mein Lehrer/meine Lehrerin lässt uns unsere Arbeitsergebnisse vorstellen.
- Schülermitbestimmung und -autonomie
 - le01: Meine Lehrerin/mein Lehrer und die Klasse diskutieren gemeinsam.
 - le02: Meine Lehrerin/mein Lehrer ermöglicht es uns, eigene Vorschläge in die Unterrichtsgestaltung einzubringen.
 - le03: Meine Lehrerin/mein Lehrer lässt uns im Unterricht Fragen stellen.
 - le04: Meine Lehrerin/mein Lehrer nimmt Beiträge und Ideen von uns Schülern/Schülerinnen in den Unterricht auf.
 - le05: Meine Lehrerin/mein Lehrer lässt uns einzelne Unterrichtsstunden mitgestalten.
 - lf01: Der Lehrer/die Lehrerin lässt uns eigene Lösungswege finden.
 - lf02: Die Gesprächsleitung haben manchmal die Schüler.
 - lf03: Bei umfangreichen Aufgaben kann ich mir die Arbeit selbst einteilen.

- lf05: Wir beurteilen uns gegenseitig.
- Unangemessenheit der Mindestanforderungen
 - uf01: Meine Lehrerin/mein Lehrer stellt im Unterricht geringe Anforderungen, sodass ich mich langweile.
 - uf02: Meine Lehrerin/mein Lehrer geht im Stoff so langsam voran, dass ich keine Lust habe, dem Unterricht zu folgen.
 - uf03: Meine Lehrerin/mein Lehrer stellt so einfache Aufgaben, dass ich gar nicht wirklich nachdenken muss.
- Unangemessenheit der Leistungserwartungen
 - ue01: Meine Lehrerin/mein Lehrer stellt oft so schwere Aufgaben, dass ich sie nicht lösen kann.
 - ue02: Meine Lehrerin/mein Lehrer gestaltet den Unterricht so schwer, dass ich nicht mitkomme.
 - ue03: Meine Lehrerin/mein Lehrer lässt uns schwierige Klassenarbeiten/Tests schreiben.
- Binnendifferenzierung
 - lb01: Bessere Schüler/-innen erhalten schwierigere Aufgaben.
 - lb02: Wir arbeiten je nach unserem Können an unterschiedlich schweren Aufgaben.
 - lb03: Arbeitsgruppen werden nach dem Können der Schüler/-innen zusammen gestellt.
- Klassen- und Zeitmanagement
 - ud01: Meine Lehrerin/mein Lehrer hat die Klasse im Griff.
 - ud02: Meine Lehrerin/mein Lehrer schafft es, für Ruhe und Ordnung zu sorgen.
 - ud03: Meine Lehrerin/mein Lehrer fordert, dass sich alle Schüler/Schülerinnen aktiv am Unterricht beteiligen.

- ud04: Meine Lehrerin/mein Lehrer kann sich in der Klasse nicht durchsetzen (**recodiert**).
- ud05: Meine Lehrerin/mein Lehrer legt Wert auf einen pünktlichen Beginn des Unterrichts.

Unterrichtsnutzung:

- Unterrichtliches Selbstengagement
 - mc01: Meine Hausaufgaben erledige ich in der Regel sorgfältig.
 - mc02: Im Unterricht lasse ich mich leicht ablenken (**recodiert**).
 - mc03: Im Unterricht arbeite ich meist konzentriert.
 - mc04: Auf Klausuren/Tests bereite ich mich in der Regel gut vor.

Schülervoraussetzungen:

- Selbsteinschätzung zur Leistungsfähigkeit
 - mb01: Ich gehöre zu den leistungsstarken Schülern/-innen in meiner Klasse.
 - mb02: In Klassenarbeiten schreibe ich häufig gute Noten.
 - mb03: Lernen fällt mir leicht.
 - mb04: Ich brauche immer viel Zeit, um Lerninhalte zu verstehen (**recodiert**).
- Fachbezogenes Interesse
 - md01: Die in diesem Unterrichtsfach behandelten Inhalte machen mir Spaß.
 - md02: Im Unterricht in diesem Fach frage ich mich oft, was ich hier soll (**recodiert**).
 - md03: Über die Inhalte dieses Fachs würde ich gerne noch mehr erfahren.

- md04: Die Inhalte in diesem Fach finde ich interessant.

Kontext:

- Einstellung zur Schule (Schulbeurteilung)
 - ga01: Mit meiner Schule bin ich sehr zufrieden.
 - ga02: Wenn ich die Wahl hätte, würde ich wieder auf diese Schule gehen.
 - ga03: Wenn mein bester Freund/meine beste Freundin fragt, würde ich diese Schule empfehlen.

Unterrichtswirksamkeit:

- Wahrgenommener Lernzuwachs
 - za01: Wenn ich auf das letzte Halbjahr zurück blicke, habe ich viel dazu gelernt.
 - za02: Bei diesem Lehrer/bei dieser Lehrerin habe ich wirklich viel gelernt.
 - za03: Im Unterricht habe ich gelernt, die verschiedenen Fachgebiete zu verknüpfen.
 - za04: Ich glaube, dass ich das, was ich im Unterricht gelernt habe, später gut verwenden kann.

Anhang C – Modifizierte Item-Konstrukt-Zuordnung auf Basis der KFA-Ergebnisse

Die auf Basis der KFA-Ergebnisse unterstellte Item-Konstrukt-Zuordnung ist folgende:

Lehrerpersönlichkeit:

- *Fachkompetenz*: ui01, ui02 und ui03.
- *Unterrichtsführung* und *Strukturiertheit*: ua01, ua04, ua05, ua06 und ua07.
- Diagnostische Kompetenz (*Diagnostik*): ug01, ug02 und ug03.
- Lehrer-Schüler-*Sozialklima*: uh02, uh03 und uh04.

Unterrichtsangebot:

- Thematische *Verknüpfung*: ub01, ub03, ub04 und ub05.
- *Interessantheit*: uc01, uc02, uc03 und uc04.
- Transparenz und *Feedback*: ma01, ma02, ma03 und ma04.
- Unterstützende *Materialien*: la01, la02, la03 und la04.

- Schüleraktivierender Unterricht (*Schüleraktivierung*): ld01, ld02, ld03 und ld05.
- *Schülermitbestimmung* und *-autonomie*: le02, le04, le05 und lf01.
- Unangemessenheit der *Mindestanforderungen*: uf01, uf02 und uf03.
- Unangemessenheit der *Leistungserwartungen*: ue01 und ue02.
- *Binnendifferenzierung*: lb01, lb02 und lb03.
- Klassen- und *Zeitmanagement*: ud01, ud02, ud04r und ud05.

Unterrichtsnutzung:

- Unterrichtliches *Selbstengagement*: mc01, mc03 und mc04.

Schülervoraussetzungen:

- Selbsteinschätzung zur *Leistungsfähigkeit*: mb01, mb03 und mb04r.
- Fachbezogenes *Interesse*: md01, md02r, md03 und md04.

Kontext:

- Einstellung zur Schule (*Schulbeurteilung*): ga01, ga02 und ga03.

Unterrichtswirksamkeit:

- Wahrgenommener *Lernzuwachs*: za01, za02 und za04.

Erklärung zur Urheberschaft

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Simon Deml

Berlin, 30. November 2009